Microfluidics

Microfluidics

Sananda Chatterjee

wisdom press
NEW DELHI

Microfluidics
by Sananda Chatterjee

This edition published by Wisdom Press,
Murari Lal Street, Ansari Road, Daryaganj,
New Delhi-110002

ISBN: 978-93-81052-75-4

Edition: 2012

Printed in India by H.S. Offset Printers, Delhi-110002.

Wisdom Press
Editorial Office: 116-A, South Anarkali, Delhi-110051.
Ph.: 011-22415687
Sales & Marketing: 4378/4-B, Murari Lal Street, Ansari Road,
Daryaganj, New Delhi-110002.
Ph.: 011-23281685, Fax: +91-11-23270680
e-mail : wisdompress@ymail.com

Preface

Microfluidics is the most modern and has become one of the most powerful subjects in the futuristic world.

In this book I have introduced different microfluidic devices and systems and also have tried to explain the different properties of microfluidic systems.

I have also introduced two new subjects 'Semiotics' and 'Flow Chemistry' is this book.

It is to the readers who will find this book for their work

Contents

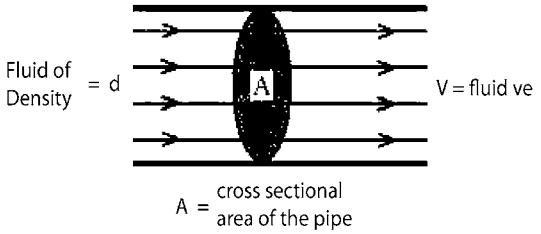
Chapter: One

Fluid Mechanics and Flow Chemistry

In this chapter we will discuss the properties of a moving fluid and what happens to objects that travel in a fluid. Two types of fluid flow will be considered: laminar flow and turbulent flow. In laminar flow the particles in the fluid follow streamlines, and the motion of particles in the fluid is predictable. If the flow rate is very large, or if objects obstruct the flow, the fluid starts to swirl in an erratic motion. No longer can one predict the exact path a particle on the fluid will follow. This region of constantly changing flow lines is said to consist of turbulent flow.

A. The Equation of Continuity—Conservation of Mass

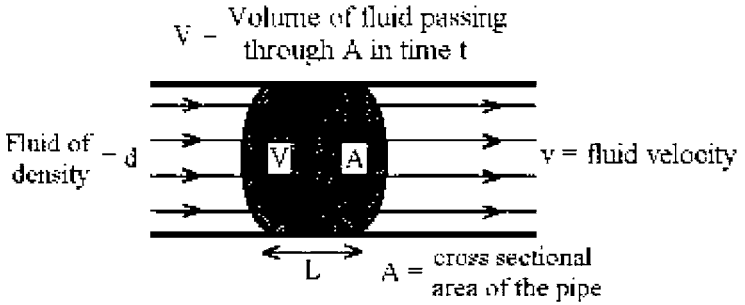
We begin our discussion of fluid dynamics by examining the simplest case of fluid flow: laminar flow with constant velocity in a uniform pipe. In the diagram below a fluid of density d flows through a uniform pipe of Area A with velocity v.



One property of interest is the amount of mass passing through area A over unit time. Since this is simply the rate at which mass passes through the pipe, we can immediately write

$$\textit{Rate at which mass flows} = \frac{m}{t}$$

To find what determines this flow rate examine the following diagram. Let us assume that in time t all mass contained in the shaded volume V will pass through area A. Let the length of this cylindrical section of fluid be L.



Now the flow rate is simply m/t. We can use the expression for density to write m = dV. This gives the relation

$$\textit{Rate at which mass flows} = \frac{m}{t} = \frac{dV}{t}$$

Now we know that the volume of water in this cylinder is

simply the product of the length of the cylinder and its base area, V = AL. So we can write

$$\textit{Rate at which mass flows} = \frac{m}{t} = \frac{dV}{t} = \frac{dAL}{t} = dA\left(\frac{L}{t}\right)$$

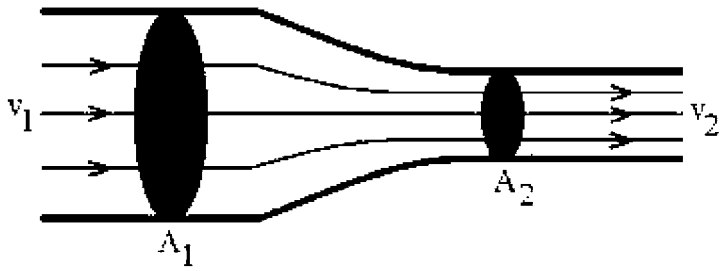
But the quantity L/t is simply the rate at which distance is covered by the fluid, that is, the fluid's velocity.

$$v = \frac{L}{t}$$

So we have an expression for the rate at which mass flows in terms of the velocity of fluid flow, density of the fluid and area of the pipe in which the fluid is flowing. This result is very reasonable.

$$\textit{Rate at which mass flows} = \frac{m}{t} = dAv \qquad (9.1)$$

We now complicate our analysis of fluid flow by examining what happens to the fluid if the size of the tubing through which it flows changes. We will allow the change to be gradual and continuous so that laminas flow is maintained. Consider the following diagram which shows the pipe slowly constricting from area A1 to area A2. From practical experience we know that the velocity of fluid through the small area is larger than the velocity of the fluid through the large area.



Many of us have heard the expression "still water runs deep." This phenomenon can be explained and quantified by examining the flow rate of mass through the tubing. Because no fluid can leave through the walls and there are no "sources" or "sinks" wherein the fluid can be created or destroyed, the mass crossing each section of the tube per unit time must be the same. This is simply the principle of *conservation of mass*. This principle is embodied in *the equation of continuity*.

$$Flow\ rate\ through\ A_1 = Flow\ rate\ through\ A_2$$

$$d_1 A_1 v_1 = d_2 A_2 v_2$$

or

$$dAv = Constant \quad (9.2)$$

This equation expresses the law of conservation of mass in fluid dynamics.

If fluid is incompressible, as will be the case with all examples considered here, then the density is constant ($d_1 = d_2$), and Eq. 9.2 takes on simpler form

$$A_1 v_1 = A_2 v_2$$

or

$$Av = Constant$$

Example: Water flows through a 1 inch diameter hose with a speed of 2 ft/sec. Find the speed of water through the nozzle of the diameter is reduced to 1/8 inch.

We use the principle of conservation of mass to solve this problem.

Av = constant

Reducing the diameter of the hose will reduce the area. Consequently the velocity must be increases by the same factor that the area is decreased. We must find by what amount the area is decreased. For a circle

Area = πr^2

where r is the radius. In this problem the diameter is reduced by a factor of eight. Subsequently, the radius is also reduced by a factor of eight. But the area is reduced by a factor of 64. This results in an increase in velocity by a factor of 64.

$$A_1 v_1 = A_2 v_2$$

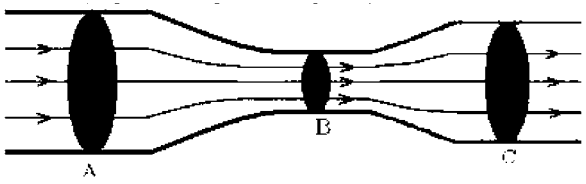
$$v_2 = \frac{A_1 v_1}{A_2} = \frac{A_1}{A_2} v_1 = 64(2\ m/\sec) = 128\ m/\sec$$

Note: Streamlines are very convenient for representing fluid flow. Notice in particular, that closely spaced streamlines indicate regions where the velocity of the fluid is great.

B. Bernoulli's Equation – Conservation of Energy

Let us continue to observe what happens to a fluid as it flows through a pipe of varying area. We have already determined that if the flow is laminar and the fluid is incompressible then the product Av is constant. Now use Newton's second law of motion and consider the pressure acting on a flowing fluid. Let us begin by considering the following question.

Question: In which region, A B, or C, in the figure below would you expect the pressure on the walls of the pipe to be the greatest? (Region A > region C > region B)



The law of conservation of mass tells us that the velocity is greatest in region B and smallest in region A. In going from a wide area to a narrow area the velocity increases. This represents an acceleration of the fluid. Acceleration requires

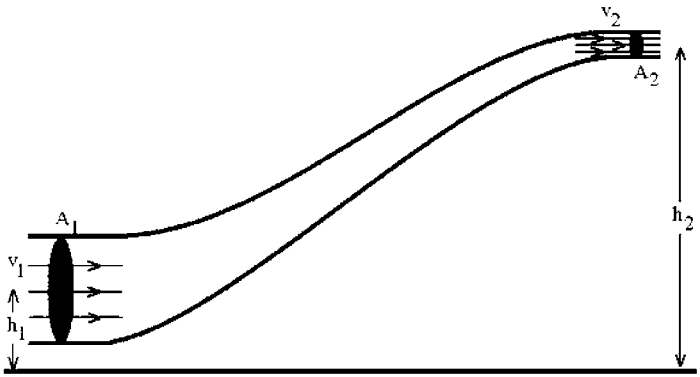
an unbalanced force, which in this case is supplied by the pressure on the fluid by the walls. Thus the pressure in region A must be greater than the pressure in region B to accelerate the fluid. Conversely, the pressure must be larger in region C to accomplish this deceleration.

In the previous discussion we found a qualitative relationship between pressure and velocity in a fluid flow. We now seek to quantify this relation. In addition, we will generalize the problem by allowing the elevation of the fluid to change. Consider the flow of fluid on the pipe shown below.

Let us examine the energy of the fluid as it flows. Because the fluid is moving it has kinetic energy (KE = ½ mv2). In addition, because of its position in the earth's gravitational filed it possesses potential energy (PE = mgh). Finally there is a force on the fluid due to the pressure (P = F/A), and this force moves the fluid a certain distance. Thus work is being done on the fluid. If there are no frictional losses in the system we can apply the law of conservation of energy.

Work in + Potential energy at the bottom + kinetic energy and the bottom =

Work out + Potential energy at the top + Kinetic energy at the top.



We would like to write this conservation law in terms of the quantities which are related to fluid flow (pressure, density, velocity, etc.). This law, written in terms of fluid variables, is called Bernoulli's equation. Bernoulli's equation states that at any point in the channel of a flowing fluid the following relationship holds.

$$P + dgh + \tfrac{1}{2} dv^2 = \text{constant} \qquad (9.4)$$

Here P is the pressure, h is the height, v is the velocity, and d is the density at any point in the flow channel.

If the elevation of the fluid remains constant, or if the change in elevation is small enough to not change the gravitational potential energy of the fluid appreciably, then the potential energy term can be ignored. We then have

$$P + \tfrac{1}{2} dv^2 = \text{constant}$$

This relationship will apply to all the problems and physical phenomena we will examine in this chapter.

Examples:

1. Water (density = 1000 kg/m^3) flows through a hose with a velocity of 1 m/sec. As it leaves the nozzle the constricted area increases the velocity to 20 m/sec. The pressure on the water as it leaves is atmospheric pressure (1 Atm = 100,000 N/m2). What is the pressure on the water in the hose? Express the answer in N/m2 and Atm.

   First let us list the information we are given.

   Inside: v_1 = 1 m/sec

   P_1 = ?

   Outside: v_2 = 20 m/sec

   P2 = 100,000 N/m^2

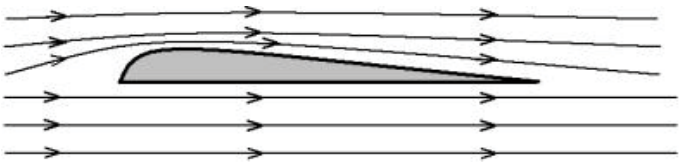
   Density of water = d = 1000 kg/m^3

Notice that care has been taken to put all values in mks units. We can now use Eq. 9.5 to find the unknown pressure P_1.

$$P_1 + \tfrac{1}{2}dv_1^2 = P_2 + \tfrac{1}{2}dv_2^2$$
$$P_1 = P_2 + \tfrac{1}{2}dv_2^2 - \tfrac{1}{2}dv_1^2$$
$$P_1 = 100{,}000\ N/m^2 + \tfrac{1}{2}(1000\ kg/m^3)(20\ m/\text{sec})^2 - \tfrac{1}{2}(1000\ kg/m^3)(1\ m/\text{sec})^2$$
$$P_1 = 100{,}000\ N/m^2 + 200{,}000\ N/m^2 - 500\ N/m^2$$
$$P_1 = 299{,}500\ N/m^2$$

Express the result in atmospheres.

2. Bernoulli's equation can also be used to show how the design of an airplane wing results in an upward lift. The flow of air around an airplane wing is illustrated below. In this case you will notice that the air is traveling faster on the upper side of the wing than on the lower.



As a result the pressure will be greater on the bottom of the wing, and the wing will be forced upward. Let us consider an airplane wing where the flow of air (density = 1.3 kg/m3) is 250 m/sec over the top of the wing and 220 m/sec over the bottom.

a. Calculate the pressure difference ($P_1 - P_2$) between the bottom and the top wing.

   Begin by listing the information given and checking that the units are consistent.

   Air velocity on bottom = v_1 = 220 m/sec

   Air velocity on top = v_2 = 250 m/sec

Air density = 1.3 kg/m^3

Pressure difference $P_1 - P_2$ = ?

Now use Bernoulli's equation

$$P_1 + \tfrac{1}{2}dv_1^2 = P_2 + \tfrac{1}{2}dv_2^2$$
$$P_1 - P_2 = \tfrac{1}{2}dv_2^2 - \tfrac{1}{2}dv_1^2$$
$$P_1 - P_2 = \tfrac{1}{2}(1.3\,kg/m^3)(250\,m/\sec)^2 - \tfrac{1}{2}(1.3\,kg/m^3)(220\,m/\sec)^2$$
$$P_1 - P_2 = 9165\,N/m^2$$

b. If the area of the two wings of the airplane is 10 m^2 what is the upward force?

The result of the first part of this example told us that the net upward pressure is 9165 N/m^2. Using the definition of pressure, P = F/A, we can write

F = PA = 9165 N/m^2 x 10 m^2 = 91,650 N

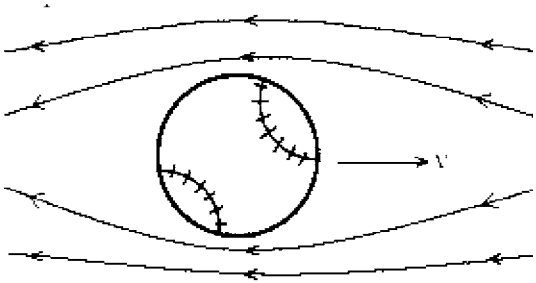
If we convert this force to pounds we obtain an upward force of approximately 20,000 lbs. This means that this air foil in question is capable of supporting a 10 ton airplane in level flight.

C. The Magnus Effect

Through our experience in sports, we all know that the path of a ball if flight is altered when the ball spins. We are perhaps most familiar with this phenomena in baseball. When a baseball pitcher throws the ball he releases it with spin. Depending upon the orientation of this spin the ball will "break" in some direction. The term "curve" and "screwball" refer to the direction in which a pitched ball "breaks". The curving of balls in flight is also frequently observed in lesser sports such as golf, tennis, and soccer.

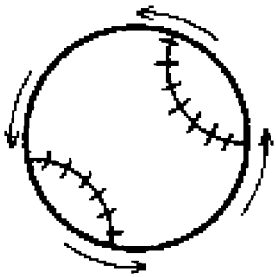
The phenomenon of spinning balls curving in flight is called the *Magnus Effect,* after the German scientist who first studied it. The Magnus Effect is described with Bernoulli's equation. Examine the flow of air about a baseball under three different conditions.

Ball moving with no spin



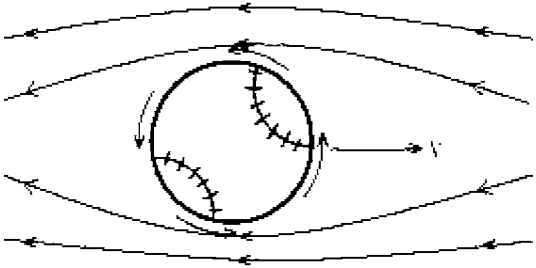
The air flow is the same on both sides. Now consider a stationary ball that is spinning counterclockwise.

Spinning Ball Without Froward Motion



As the ball spins air is dragged around with the ball. We notice here that the amount of air dragged around the spinning ball depends on the smoothness of the ball. The seams on a baseball will enhance this air flow. The next picture illustrates what happens if the spinning ball is also moving to the right.

Ball Moving to Right While Spinning Counterclockwise



The motion of air past the ball is most rapid above the ball and lowest below the ball. According to Bernoulli's equation,

$P + \frac{1}{2} dv^2 = \text{constant}$

When the fluid velocity is large, the pressure is small; and when the fluid velocity is small, the pressure is large. So the air pressure is least above the ball and greatest below. The ball will experience a net force toward the top of the page and therefore follow a curved path. A baseball thrown in this manner is called a rising fast ball.

D. Fluid Resistance

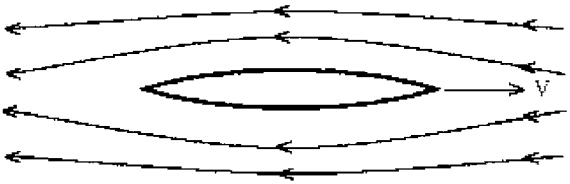
In Chapter 3 we examined a problem where a skier raced straight down a one kilometer hill inclined at an angle of 30 degrees. His velocity at the bottom of the hill was found to be an incredible 222 mph. We argued that this result was not believable, and we used this example to introduce friction, a force that opposes motion.

If the skier example is examined more closely it appears that frictional force resulting from contact between the skis and the snow is not large enough to reduce the skier's velocity to a level we would expect. This suspicion becomes more obvious when we see that the coefficient if friction between wood and snow is a relatively small value ($\mu k = 0.06$). We are

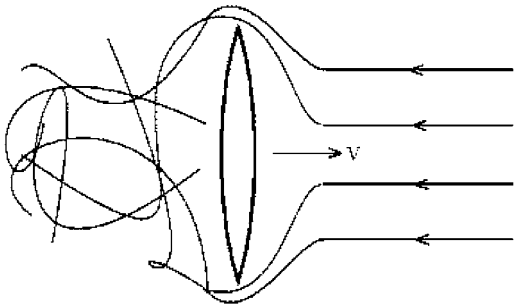
forced to conclude that other resistive forces must be acting on the skier.

With the exception of objects that move in a vacuum, we know that motion always involves a resistive medium such as air or water. Let us now investigate the phenomenon of fluid resistance. We begin by looking at the most obvious source of fluid resistance, called surface drag.

Examine the flow of air around a streamlined airfoil that bears a striking resemblance to a flying saucer. The air follows streamlines of simple laminar flow as shown below. The air is disturbed very little by this motion.



Now examine what happens when the airfoil is rotated by 90 degrees and moves to the left.



In this orientation the wing is no longer a streamlined airfoil. Instead of laminar flow around the wing, the result is turbulent flow. The motion of the wing has caused violent agitation of the air through which it passes. In the process, the wing is doing work by moving great amounts of air around its

surface. This work is done at the expense of the wing's kinetic energy. The result is a force that slows the motion of the wing. This type of resistive force is called surface drag.

Surface drag is influenced by the following factors:

Cross sectional area of the object.

Smoothness of the object

Velocity of the object.

Flow characteristics of the fluid.

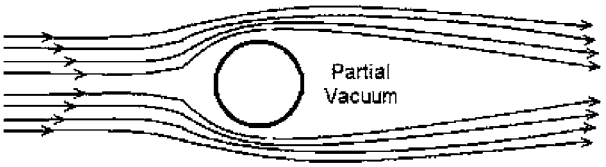
The cross sectional area determines the amount of fluid that is disturbed as the object flows through it. Clearly, in order to reduce surface drag it is important to minimize cross sectional area. This principle is demonstrated by the human body positions used in sports such as ski jumping, speed skating, and downhill racing.

During the America's Cup races the boats were removed from the water every night and their hulls were scrubbed and polished. This was done to remove any barnacles, roughness or protrusion from the hull that could increase turbulence, hence increasing surface drag.

The speed at which an object moves through a fluid also affects the amount of turbulence. You can demonstrate this by simply moving your hand in water. If you move your hand slowly, little resistance is experienced. However, when your hand moves rapidly through the water, you feel a very strong force resisting its motion.

The last factor influencing surface drag depends on the fluid through which a body moves. As the body moves it displaces fluid around it. The measure of how much force is required to slide the fluid around the body is called the viscosity of the fluid. Fluids that do nor flow easily, such as syrup, have a large viscosity. Substances (like water and air) that flow easily have small viscosity.

Secondary to surface drag, there is another type of fluid resistance called form drag. To examine this phenomenon let us look at a spinless ball in flight with laminar flow. The laminar flow of air around the ball is shown below. As the ball moves through the air a partial vacuum is created behind the ball. As a result the air pressure is greater in front of the ball. The unbalanced pressures result in unbalanced forces that slow the ball down.



From drag can be reduced by streamlining a body. This streamlining eliminates the partial vacuum behind the body. Airplane wings are streamlined in such a manner so as to keep form drag at a minimum.

It is obvious that objects such as baseballs and golf balls cannot be streamlined. Form drag will have a large effect on their motion if it is not reduced. This is accomplished by the rotation of the balls. As a ball spins it drags air around its surface. This air will help to fill the partial vacuum behind the ball. Subsequently, the unbalanced pressures on the ball will be reduced. The dimples on a golf ball and threads on a baseball will increase the amount of air dragged around them, thus reducing the pressure differences even more.

Summary

Laminar Flow
Turbulent Flow
Flow Rate
The Equation of Continuity
Bernoulli's Equation
Magnus Effect
Surface Drag
Viscosity
Form Drag

Problems

1. Water is flowing at a rate of 3 m/sec in a horizontal pipe under a pressure of 200,000 N/m^2. If the pipe narrows to half its original diameter
   a. What is the new speed of flow?
   b. What is the new pressure?
   c. How does the flow rate through the narrow section compare with the flow rate through the wider section?
2. Water flows through a 2 cm diameter hose with a speed of 0.5 m/sec. Find the speed of water through the nozzle if the diameter is reduced to 0.4 cm.
3. Blood flows through the aorta at a rate of 5 liter/min. Determine the flow rate in gm/sec. (The density of blood is 1.06 gm/cm^3).
4. Compute the speed of blood flow in the aorta of radius 1 cm (Area = 3.14 cm^2), if the flow rare is 5 liter/min (88 gm/sec).
5. Hardening of the arteries results in a constriction of the arteries. Blood flows with a speed of 26.5 cm/sec through an aorta of radius 1 cm. (Area = 3.14 cm^2). If the flow of blood is to remain constant, calculate the velocity of blood in a region where the aorta has constricted to a radius of 0.8 cm (Area = 2.01 cm^2) dues to hardening of the arteries.
6. If the flow of blood is to remain constant through the aorta the velocity of blood must increase from 26.5 cm/sec to 40 cm/sec as the radius decreases from 1 cm to 0.8 cm. Calculate the pressure difference between these two areas of the aorta in N/m^2. (Density of blood = 1060 kg/m^3).

7. The following questions are simply applications of Bernoulli's principle.
    a. Why is the draft in a fireplace better on a windy day?
    b. What keeps a Frisbee in flight?
    c. Why does you car lurch toward an oncoming truck as it passes by?
    d. Why does the wind increase the size of water waves?
8. The air above the wing of a plane travels at a speed of 320 m/sec; below the wing the speed of air is 300 m/sec. The density of air is 1.3 kg/m^3.
    a. Calculate the pressure difference between the bottom and the top of the wing.
    b. If the wing area of the plane is 8 m^2 calculate the upward force (lift) on the plane.
9. The luge is a small sled in which the racer lies on his back and rides down the course feet first. This is a very dangerous sport and participants always wear helmets. In 1983 the East German luge team introduced helmets which were cone shaped. Honestly, it looked like they had just stepped off the set of "Saturday Night Live." What competitive advantage would these helmet give s luge racer?
10. These helmets were quickly outlawed because they were considered as dangerous as no helmet at all. Why were these helmets considered to be so dangerous?

Laminar flow

Laminar flow, sometimes known as streamline flow, occurs when a fluid flows in parallel layers, with no disruption between

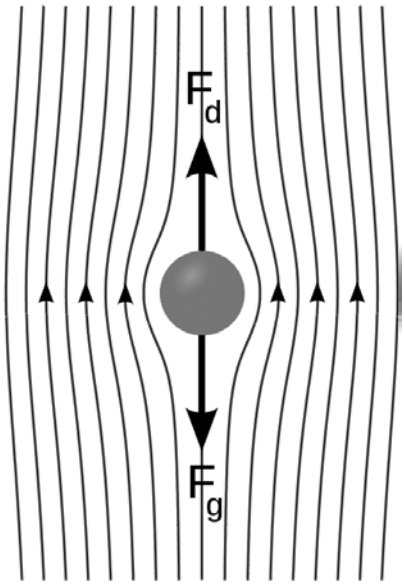


Fig. 1—Showing An object moving through a gas or liquid experiences a force in direction opposite to its motion. Terminal velocity is achieved when the drag force is equal in magnitude but opposite in direction to the force propelling the object. Shown is a sphere in Stokes flow, at very low Reynolds number

the layers. At low velocities the fluid tends to flow without lateral mixing, and adjacent layers slide past one another like playing cards. There are no cross currents perpendicular to the direction of flow, nor eddies or swirls of fluids. In laminar flow the motion of the particles of fluid is very orderly with all particles moving in straight lines parallel to the pipe walls. In fluid dynamics, laminar flow is a flow regime characterized by high momentum diffusion and low momentum convection. When a fluid is flowing through a closed channel such as a pipe or between two flat plates, either two types of flow may occur depending on the velocity of the fluid: laminar flow or turbulent flow. Laminar flow is the opposite of turbulent flow which occurs at higher velocities where eddies or small packets of fluid particles form leading to lateral mixing. In nonscientific terms laminar flow is "smooth", while turbulent flow is "rough." The type of flow occurring in a fluid in a channel is important in fluid dynamics problems. The dimensionless Reynolds number is an important parameter in the equations that describe whether flow conditions lead to laminar or turbulent flow. In the case of flow through a straight pipe with a circular cross-section,

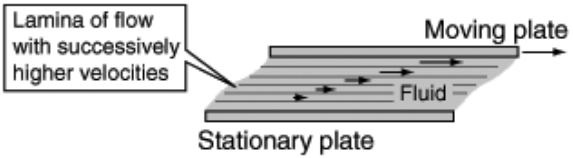


Fig. 2—Showing In the case of a moving plate in a liquid, it is found that there is a layer or lamina which moves with the plate, and a layer which is essentially stationary if it is next to

Reynolds numbers of less than 2100 are generally considered to be of a laminar type; however, the Reynolds number upon which laminar flows become turbulent is dependent upon the flow geometry. When the Reynolds number is much less than 1, Creeping motion or Stokes flow occurs. This is an extreme case of laminar flow where viscous (friction) effects are much greater than inertial forces. The common application of laminar flow would be in the smooth flow of a viscous liquid through a tube or pipe. In that case, the velocity of flow varies from zero at the walls to a maximum along the centerline of the vessel. The flow profile of laminar flow in a tube can be calculated by dividing the flow into thin cylindrical elements and applying the viscous force to them. For example, consider the flow of air over an aircraft wing. The boundary layer is a very thin sheet of air lying over the surface of the wing (and all other surfaces of the aircraft). Because air has viscosity, this layer of air tends to adhere to the wing. As the wing moves forward through the air, the boundary layer at first flows smoothly over the streamlined shape of the airfoil. Here the flow is called laminar and the

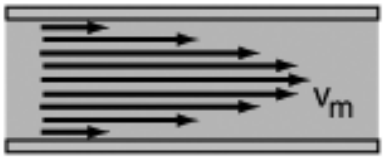


Fig 3—Showing The streamlines associated with laminar flow are resemble a deck of cards. This flow profile of a fluid in a pipe shows that the fluid acts in layers and slides over one another.

boundary layer is a laminar layer. Prandtl applied the concept of the laminar boundary layer to airfoils in 1904.

Flow Chemistry

Flow chemistry, sometimes referred to as plug flow, microchemistry or continuous flow chemistry has emerged as a new and productive technology for R&D chemists. It offers the chemist a new tool to speed discovery and development.

Flow chemistry has many useful benefits over conventional batch techniques. It is a complementary technology, and chemists can use it as necessary without major change to their work.

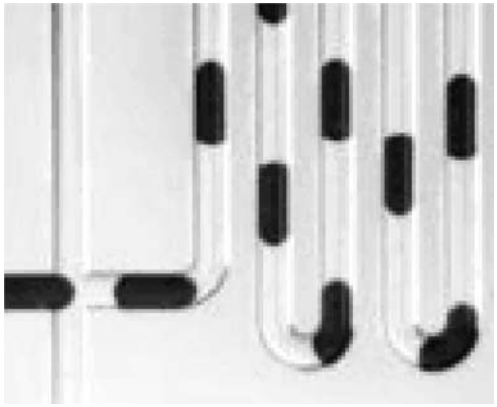


Fig 4—Showing Bubble flow in a flow reactor

This section provides information about the application of flow chemistry. Please use the navigation to the right to access more specialist flow chemistry information.

Syrris have designed and manufacture two well established flow chemistry systems (Africa and FRX) which can be used to access the advantages of flow chemistry, listed below.

Syrris' experience as the world's longest established

provider of lab scale flow chemistry systems has been used to develop Asia, a revolutionary range of advanced flow chemistry products. Asia is available from February 2011.Click here to find out more about Asia

Advantages of Flow Reactors Include

Control

The system gives excellent, repeatable control of reaction conditions, *i.e.* time, temperature, equivalents of reagents and mixing, and allows exothermic reactions to be performed without the need for cryogenics.

Fig 5—Showing the flow of Mixing of liquids in a Syrris glass microreactor

Increased Reaction Rate

By easily pressurising the system (e.g. to 300psi), reactions can be superheated to give reaction rates hundreds of times faster than in reflux. This gives microwave-like rate enhancement without the problems associated with scaling up microwave-based reactions.

Integrated Synthesis and Analysis

Reactions flow immediately from the reactor to the HPLC via a sampler and dilutor that takes a 5ul sample from the reaction mixture and dilutes it before flowing it onto the HPLC.

Automation

Fig 6—Showing the Flow reaction showing mixing in a microreactor

The reaction and analysis conditions are simply defined

in the software and the system runs a variety of reaction conditions unattended.

Efficiency

The system operates with low raw materials consumption when working in method development mode e.g. 100ul per reaction.

Range of Use

The system can be used for synthesis, reaction optimisation or for libraries, operating from 1mg per experiment to 200g per day. For example to optimise a reaction, a number of reactions can be performed using a structured exploration of parameters such as time, temperature and stoichiometry.

Here you can find some of the published flow chemistry performed on Syrris flow chemistry systems since 2006. Including 'Fully Automated Continuous Flow Synthesis of Highly Functionalized Imidazo[1,2-a] Heterocycles'

Africa and FRX give good process scalability from discovery through to development and ultimately to manufacture.

Closed Loop

The system has potential for closed loop reaction optimisation in the medium term and closed loop drug discovery in the long term.

Fig 7—Showing the Flow star close up system in authors lab

Hagen–Poiseuille Equation

In fluid dynamics, the *Hagen–Poiseuille equation* is a physical law that gives the pressure drop in a fluid flowing through a long cylindrical pipe. The assumptions of the equation are that the flow is laminar viscous and incompressible and the flow is through a constant circular cross-section that is substantially longer than its diameter. The equation is also known as the *Hagen–Poiseuille law*, *Poiseuille law* and *Poiseuille equation*.

History

It was developed independently by Gotthilf Heinrich Ludwig Hagen (1797-1884) and Jean Louis Marie Poiseuille. Poiseuille's law was experimentally derived in 1838 and formulated and published in 1840 and 1846 by Jean Louis Marie Poiseuille (1797–1869). Hagen did his experiments in 1839The fluid flow will be turbulent for velocities and pipe diameters above a threshold, leading to larger pressure drops than would be expected according to the Hagen–Poiseuille equation.

Equation

Standard Fluid Dynamics Notation

In standard fluid dynamics notation:

$$\Delta P = \frac{8\mu LQ}{\pi r^4}$$

or

$$\Delta P = \frac{128\mu LQ}{\pi d^4}$$

where:

ΔP is the pressure drop

L is the length of pipe

μ is the dynamic viscosity

Q is the volumetric flow rate

r is the radius

d is the diameter

π is the mathematical constant

Physics Notation

$$\Phi = \frac{dV}{dt} = v\pi R^2 = \frac{\pi R^4}{8\eta}\left(\frac{-\Delta P}{\Delta x}\right) = \frac{\pi R^4}{8\eta}\frac{|\Delta P|}{L}$$

where:

Φ is the volumetric flow rate

V is a volume of the liquid poured (cubic meters)

t is the time (seconds)

v is mean fluid velocity along the length of the tube (meters/second)

x is a distance in direction of flow (meters)

R is the internal radius of the tube (meters)

ΔP is the pressure difference between the two ends (pascals)

η is the dynamic fluid viscosity (pascal-second (Pa•s)),

L is the total length of the tube in the x direction (meters).

Relation to Darcy–Weisbach

This result is also a solution to the phenomenological Darcy–Weisbach equation in the field of hydraulics, given a relationship for the friction factor in terms of the Reynolds number:

$$\Lambda = \frac{64}{\mathrm{Re}}\,, \qquad \mathrm{Re} = \frac{2\rho v r}{\eta}\,,$$

where *Re* is the Reynolds number and ρ fluid density. In this form the law approximates the *Darcy friction factor*, the *energy* (*head*) *loss factor*, *friction loss factor* or *Darcy* (*friction*) *factor* Λ in the laminar flow at very low velocities in cylindrical tube. The theoretical derivation of a slightly different form of the law was made independently by Wiedman in 1856 and Neumann and E. Hagenbach in 1858 (1859, 1860). Hagenbach was the first who called this law the Poiseuille's law.

The law is also very important specially in hemorheology and hemodynamics, both fields of physiology.[1]

The Poiseuilles' law was later in 1891 extended to turbulent flow by L. R. Wilberforce, based on Hagenbach's work.

Derivation

Main article: Hagen–Poiseuille flow from the Navier–Stokes equations

The Hagen–Poiseuille equation can be derived from the Navier–Stokes equations.

Viscosity

The derivation of Poiseuille's law is surprisingly simple, but it requires an understanding of viscosity. When two layers of liquid in contact with each other move at different speeds, there will be a force between them. This force is proportional to the area of contact A, the velocity difference in the direction of flow Δvx/Δvy, and a proportionality constant η and is given by

$$F_{\text{viscosity, top}} = -\eta A \frac{\Delta v_x}{\Delta v_y}.$$

The negative sign is in there because we are concerned

with the faster moving liquid (top in figure), which is being slowed by the slower liquid (bottom in figure). By Newton's third law of motion, the force on the slower liquid is equal and opposite (no negative sign) to the force on the faster liquid. This equation assumes that the area of contact is so large that we can ignore any effects from the edges and that the fluids behave as Newtonian fluids.

Fig 8—Showing Two fluids moving past each other in the x direction. The liquid on top is moving faster and will be pulled in the negative direction by the bottom liquid while the bottom liquid will be pulled in the positive direction by the top liquid.

Liquid flow through a pipe

In a tube we make a basic assumption: the liquid in the center is moving fastest while the liquid touching the walls of the tube is stationary (due to friction).

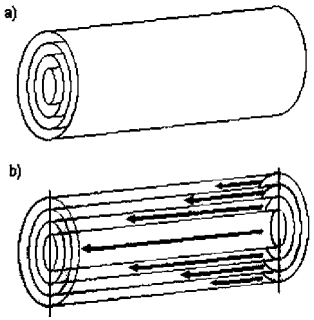


Fig 9—Showing (a) A tube showing the imaginary lamina. (b) A cross section of the tube shows the lamina moving at different speeds. Those closest to the edge of the tube are moving slowly while those near the center are moving quickly.

To simplify the situation, let's assume that there are a

bunch of circular layers (lamina) of liquid, each having a velocity determined only by their radial distance from the center of the tube.

To figure out the motion of the liquid, we need to know all forces acting on each lamina:

1. The force pushing the liquid through the tube is the change in pressure multiplied by the area: $F = -\Delta PA$. This force is in the direction of the motion of the liquid—the negative sign comes from the conventional way we define $\Delta P = P_{end} - P_{top} < 0$.
2. The pull from the faster lamina immediately closer to the center of the tube
3. The drag from the slower lamina immediately closer to the walls of the tube.

The first of these forces comes from the definition of pressure. The other two forces require us to modify the equations above that we have for viscosity. In fact, we are not modifying the equations, instead merely plugging in values specific to our problem. Let's focus on the pull from the faster lamina (#2) first.

Faster Lamina

Assume that we are figuring out the force on the lamina with radius s. From the equation above, we need to know the area of contact and the velocity gradient. Think of the lamina as a cylinder of radius s and thickness ds. The area of contact between the lamina and the faster one is simply the area of the inside of the cylinder: $A = 2\pi s\Delta x$. We don't know the exact form for the velocity of the liquid within the tube yet, but we do know (from our assumption above) that it is dependent on the radius. Therefore, the velocity gradient is the change of the velocity with respect to the change in the radius at the intersection of these two laminae. That intersection is at a

radius of s. So, considering that this force will be positive with respect to the movement of the liquid (but the derivative of the velocity is negative), the final form of the equation becomes

$$\text{scosity, fast} = -\eta 2\pi s \Delta x \left.\frac{d}{d}\right.$$

where the vertical bar and subscript s following the derivative indicates that it should be taken at a radius of s.

Slower lamina

Next let's find the force of drag from the slower lamina. We need to calculate the same values that we did for the force from the faster lamina. In this case, the area of contact is at s+ds instead of s. Also, we need to remember that this force opposes the direction of movement of the liquid and will therefore be negative (and that the derivative of the velocity is negative).

Putting it all together

To find the solution for the flow of liquid through a tube, we need to make one last assumption. There is no acceleration of liquid in the pipe, and by Newton's first law, there is no net force. If there is no net force then we can add all of the forces together to get zero

$$0 = F_{pressure} + F_{viscosity,\,\text{fast}} + F_{viscosity,\,slow}$$

or

$$F_{\text{viscosity, slow}} = \eta 2\pi (s + ds)\Delta x \left.\frac{dv}{dr}\right|_{s+ds}$$

Before we move further, we need to simplify this equation. First, to get everything happening at the same point, we need to do a Taylor series expansion of the velocity gradient, keeping

only the linear and quadratic terms (a standard mathematical trick).

$$\left.\frac{dv}{dr}\right|_{r+dr} = \left.\frac{dv}{dr}\right|_{r} + \left.\frac{d^2v}{dr^2}\right|_{r} dr$$

Let's use this relation in our equation. Also, let's use r instead of s since the lamina we chose was arbitrary and we want our expression to be valid for all laminae. Grouping like terms and dropping the vertical bar since all derivatives are assumed to be at radius *r*,

$$0 = -\Delta P 2\pi r dr + \eta 2\pi dr \Delta x \frac{dv}{dr} + \eta 2\pi r dr \Delta x \frac{d^2v}{dr^2} + \eta 2\pi (dr)^2 \Delta x \frac{d^2v}{dr^2}.$$

Finally, let's get this in the form of a differential equation, moving some terms around to make it easier to solve later, and neglecting the term quadratic in dr since this will be really small compared to the rest (another standard mathematical trick).

$$\frac{1}{\eta}\frac{\Delta P}{\Delta x} = \frac{d^2v}{dr^2} + \frac{1}{r}\frac{dv}{dr}$$

It can be seen that both sides of the equations are negative: there is a drop of pressure along the tube (left side) and both first and second derivatives of the velocity are negative (velocity has a maximum value of the center of the tube). The equation may be re-arranged to:

$$\frac{1}{\eta}\frac{\Delta P}{\Delta x} = \frac{1}{r}\frac{d}{dr}r\frac{dv}{dr}.$$

This differential equation is subject to the following boundary conditions:

v(r) = 0 at r = R -- "No-slip" Boundary Condition at the Wall

$\frac{dv}{dr} = 0$ at r = 0 -- Axial symmetry.

Axial symmetry means that the velocity v(r) is maximum at the center of the tube, therefore the first derivative $\frac{dv}{dr}$ is zero at r = 0.

The differential equation can be integrated to:

$$v(r) = \frac{1}{4\eta} r^2 \frac{\Delta P}{\Delta x} + A \ln(r) + B.$$

To find A and B, we use the boundary conditions.

First, the symmetry boundary condition indicates:

$$\frac{dv}{dr} = \frac{1}{2\eta} r \frac{\Delta P}{\Delta x} + A \frac{1}{r} = 0 \text{ at r = 0.}$$

A solution possible only if A = 0. Next the no-slip boundary condition is applied to the remaining equation:

$$v(R) = \frac{1}{4\eta} R^2 \frac{\Delta P}{\Delta x} + B = 0$$

so therefore

$$B = -\frac{1}{4\eta} R^2 \frac{\Delta P}{\Delta x}.$$

Now we have a formula for the velocity of liquid moving through the tube as a function of the distance from the center of the tube

$$v = -\frac{1}{4\eta} \frac{\Delta P}{\Delta x} (R^2 - r^2)$$

or, at the center of the tube where the liquid is moving fastest (r = 0) with R being the radius of the tube,

$$v_{max} = -\frac{1}{4\eta} \frac{\Delta P}{\Delta x} R^2.$$

Poiseuille's Law

To get the total volume that flows through the tube, we need to add up the contributions from each lamina. To calculate the flow through each lamina, we multiply the velocity (from above) and the area of the lamina.

$$\Phi(r)dr = \frac{1}{4\eta}\frac{|\Delta P|}{\Delta x}(R^2 - r^2)2\pi r dr = \frac{\pi}{2\eta}\frac{|\Delta P|}{\Delta x}(rR^2 - r^3)dr$$

Finally, we integrate over all lamina via the radius variable *r*.

$$\Phi = \frac{\pi}{2\eta}\frac{|\Delta P|}{\Delta x}\int_0^R (rR^2 - r^3)\,dr = \frac{|\Delta P|\pi R^4}{8\eta\Delta x}$$

Poiseuille's Equation for Compressible Fluids

For a compressible fluid in a tube the volumetric flow rate and the linear velocity is not constant along the tube. The flow is usually expressed at outlet pressure. As fluid is compressed or expands, work is done and the fluid is heated and cooled. This means that the flow rate depends on the heat transfer to and from the fluid. For an ideal gas in the isothermal case, where the temperature of the fluid is permitted to equilibrate with its surroundings, and when the pressure difference between ends of the pipe is small, the volumetric flow rate at the pipe outlet is given by

$$\Phi = \frac{dV}{dt} = v\pi R^2 = \frac{\pi R^4 (P_i - P_o)}{8\eta L} \times \frac{P_i + P_o}{2P_o} = \frac{\pi R^4}{16\eta L}\left(\frac{P_i^2 - P_o^2}{P_o}\right)$$

where:

P_i inlet pressure

P_o outlet pressure

L is the length of tube

η is the viscosity

R is the radius

V is the volume of the fluid at outlet pressure

v is the velocity of the fluid at outlet pressure

This is usually a good approximation when the flow velocity is less than mach 0.3

This equation can be seen as Poiseuille's law with an extra correction

$$\frac{P_i + P_o}{2} \times \frac{1}{P_o}$$

factor expressing the average pressure relative to the outlet pressure.

Electrical Circuits Analogy

Electricity was originally understood to be a kind of fluid. This hydraulic analogy is still conceptually useful.

Poiseuille's law corresponds to Ohm's law for electrical circuits (V = IR), where the pressure drop ΔP is analogous to the voltage V and volumetric flow rate Φ is analogous to the current I. Then the resistance

$$R = \frac{8\eta\Delta x}{\pi r^4}.$$

This concept is useful because the effective resistance in a tube is inversely proportional to the fourth power of the radius. This means that halving the radius of the tube increases the resistance to fluid movement by a factor of 16.

Both Ohm's law and Poiseuille's law illustrate transport phenomena.

Darcy–Weisbach Equation

In fluid dynamics, the Darcy–Weisbach equation is a phenomenological equation, which relates the head loss—or pressure loss—due to friction along a given length of pipe to the average velocity of the fluid flow. The equation is named

after Henry Darcy and Julius Weisbach. The Darcy–Weisbach equation contains a dimensionless friction factor, known as the Darcy friction factor. This is also called the Darcy–Weisbach friction factor or Moody friction factor. The Darcy friction factor is four times the Fanning friction factor, with which it should not be confused.

History

Historically this equation arose as a variant on the Prony equation; this variant was developed by Henry Darcy of France, and further refined into the form used today by Julius Weisbach of Saxony in 1845. Initially, data on the variation of f with velocity was lacking, so the Darcy–Weisbach equation was outperformed at first by the empirical Prony equation in many cases. In later years it was eschewed in many special-case situations in favor of a variety of empirical equations valid only for certain flow regimes, notably the Hazen-Williams equation or the Manning equation, most of which were significantly easier to use in calculations. However, since the advent of the calculator, ease of calculation is no longer a major issue, and so the Darcy–Weisbach equation's generality has made it the preferred one.

Head loss form

Head loss can be calculated with

$$h_f = f \cdot \frac{L}{D} \cdot \frac{V^2}{2g}$$

where

h_f is the head loss due to friction;

L is the length of the pipe;

D is the hydraulic diameter of the pipe (for a pipe of circular section, this equals the internal diameter of the pipe);

V is the average velocity of the fluid flow, equal to the

volumetric flow rate per unit cross-sectional wetted area;

g is the local acceleration due to gravity;

f is a dimensionless coefficient called the Darcy friction factor. It can be found from a Moody diagram or more precisely by solving Colebrook equation.

Pressure Loss Form

Given that the head loss h_f expresses the pressure loss Δp as the height of a column of fluid,

$$\Delta p = \rho \cdot g \cdot h_f$$

where ρ is the density of the fluid, the Darcy–Weisbach equation can also be written in terms of pressure loss:

$$\Delta p = f \cdot \frac{L}{D} \cdot \frac{\rho V^2}{2}$$

where the pressure loss due to friction Δp is a function of:

—the ratio of the length to diameter of the pipe, L/D;

—the density of the fluid, ρ;

—the mean velocity of the flow, V, as defined above;

—a (dimensionless) coefficient of laminar, or turbulent flow, f.

Since the pressure loss equation can be derived from the head loss equation by multiplying each side by ρ and g.

Darcy Friction Factor

The friction factor f or flow coefficient λ is not a constant and depends on the parameters of the pipe and the velocity of the fluid flow, but it is known to high accuracy within certain flow regimes. It may be evaluated for given conditions by the use of various empirical or theoretical relations, or it may be

See also Darcy friction factor formulae

obtained from published charts. These charts are often referred to as Moody diagrams, after L. F. Moody, and hence the factor itself is sometimes called the Moody friction factor. It is also sometimes called the Blasius friction factor, after the approximate formula he proposed. For laminar (slow) flows, it is a consequence of Poiseuille's law that $\lambda=64/Re$, where Re is the Reynolds number calculated substituting for the characteristic length the hydraulic diameter of the pipe, which equals the inside diameter for circular pipe geometries.For turbulent flow, methods for finding the friction factor f include using a diagram such as the Moody chart; or solving equations such as the Colebrook-White equation, or the Swamee-Jain equation. While the diagram and Colebrook-White equation are iterative methods, the Swamee-Jain equation allows f to be found directly for full flow in a circular pipe.

Confusion with the Fanning Friction Factor

The Darcy–Weisbach friction factor is 4 times larger than the Fanning friction factor, so attention must be paid to note which one of these is meant in any "friction factor" chart or equation being used. Of the two, the Darcy–Weisbach factor is more commonly used by civil and mechanical engineers, and the Fanning factor by chemical engineers, but care should be taken to identify the correct factor regardless of the source of the chart or formula.Most charts or tables indicate the type of friction factor, or at least provide the formula for the friction factor with laminar flow. If the formula for laminar flow is $f = 16/Re$, it's the Fanning factor, and if the formula for laminar flow is $f = 64/Re$, it's the Darcy–Weisbach factor. Which friction factor is plotted in a Moody diagram may be determined by inspection if the publisher did not include the formula described above:

1. Observe the value of the friction factor for laminar flow at a Reynolds number of 1000.
2. If the value of the friction factor is 0.064, then the Darcy friction factor is plotted in the Moody diagram. Note that the nonzero digits in 0.064 are the numerator in the formula for the laminar Darcy friction factor: f = 64/Re.
3. If the value of the friction factor is 0.016, then the Fanning friction factor is plotted in the Moody diagram. Note that the nonzero digits in 0.016 are the numerator in the formula for the laminar Fanning friction factor: f = 16/Re.

The procedure above is similar for any available Reynolds number that is an integral power of ten. It is not necessary to remember the value 1000 for this procedure – only that an integral power of ten is of interest for this purpose.

Derivation

The Darcy–Weisbach equation is a phenomenological formula obtainable by dimensional analysis.

Away from the ends of the pipe, the characteristics of the flow are independent of the position along the pipe. The key quantities are then the pressure drop along the pipe per unit length, $\Delta p/L$, and the volumetric flow rate.

The flow rate can be converted to an average velocity V by dividing by the wetted area of the flow (which equals the cross-sectional area of the pipe if the pipe is full of fluid). Pressure has dimensions of energy per unit volume. Therefore, the pressure drop between two points must be proportional to $(1/2)\rho V2$, which has the same dimensions as it resembles (see below) the expression for the kinetic energy per unit volume. We also know that pressure must be proportional to the length

of the pipe between the two points L as the pressure drop per unit length is a constant. To turn the relationship into a proportionality coefficient of dimensionless quantity we can divide by the hydraulic diameter of the pipe, D, which is also constant along the pipe. Therefore,

$$\Delta p \propto \frac{L}{D} \cdot \frac{1}{2}\rho V^2.$$

The proportionality coefficient is the dimensionless "Darcy friction factor" or "flow coefficient". This dimensionless coefficient will be a combination of geometric factors such as π, the Reynolds number and (outside the laminar regime) the relative roughness of the pipe (the ratio of the roughness height to the hydraulic diameter).

> ***Note*** *that (1/2)ρV² is not the kinetic energy of the fluid per unit volume, for the following reasons. Even in the case of laminar flow, where all the flow lines are parallel to the length of the pipe, the velocity of the fluid on the inner surface of the pipe is zero due to viscosity, and the velocity in the center of the pipe must therefore be larger than the average velocity obtained by dividing the volumetric flow rate by the wet area. The average kinetic energy then involves the mean-square velocity, which always exceeds the square of the mean velocity. In the case of turbulent flow, the fluid acquires random velocity components in all directions, including perpendicular to the length of the pipe, and thus turbulence contributes to the kinetic energy per unit volume but not to the average lengthwise velocity of the fluid.*

Practical applications

In hydraulic engineering applications, it is often desirable to express the head loss in terms of volumetric flow rate in the

pipe. For this, it is necessary to substitute the following into the original head loss form of the Darcy-Weisbach equation

$$V^2 = \frac{Q^2}{A_w^2}$$

where

- V is, as above, the average velocity of the fluid flow, equal to the volumetric flow rate per unit cross-sectional wetted area;
- Q is the volumetric flow rate;
- A_w is the cross-sectional wetted area;

For the general case of an arbitrarily-full pipe, the value of A_w will not be immediately known, being an implicit function of pipe slope, cross-sectional shape, flow rate and other variables. If, however, the pipe is assumed to be full flowing and of circular cross-section, as is common in practical scenarios, then

$$A_w^2 = (\frac{\pi D}{4})^2 = \frac{\pi^2 D^2}{16}$$

where D is the diameter of the pipe

Substuting these results into the original formulation yields the final equation for head loss in terms of volumetic flow rate in a full-flowing circular pipe

$$h_f = \frac{8fLQ^2}{g\pi^2 D^3}$$

where all symbols are defined as above.

Darcy Friction Factor Formulae

In fluid dynamics, the Darcy friction factor formulae are equations—based on experimental data and theory—for the Darcy friction factor. The Darcy friction factor is a dimensionless quantity used in the Darcy–Weisbach equation,

for the description of friction losses in pipe flow as well as open channel flow. It is also known as the Darcy–Weisbach friction factor or Moody friction factor and is four times larger than the Fanning friction factor.

Flow Regime

Which friction factor formula may be applicable depends upon the type of flow that exists:

- Laminar flow
- Transition between laminar and turbulent flow
- Fully turbulent flow in smooth conduits
- Fully turbulent flow in rough conduits
- Free surface flow.

Laminar Flow

The Darcy friction factor for laminar flow (Reynolds number less than 2000) is given by the following formula:

$$f = \frac{64}{\mathrm{Re}}$$

where:

- f is the Darcy friction factor
- Re is the Reynolds number.

Transition Flow

Transition (neither fully-laminar nor fully-turbulent) flow occurs in the range of Reynolds numbers between 2300 and 4000. The value of the Darcy friction factor may be subject to large uncertainties in this flow regime.

Turbulent Flow in Smooth Conduits

Empirical correlations exist for this flow regime. Such correlations are included in the ASHRAE *Handbook of Fundamentals*.

Turbulent Flow in Rough Conduits

The Darcy friction factor for fully turbulent flow (Reynolds number greater than 4000) in rough conduits is given by the Colebrook equation.

Free Surface Flow

The last formula in the Colebrook equation section of this article is for free surface flow. The approximations elsewhere in this article are not applicable for this type of flow.] Choosing a formula

Before choosing a formula it is worth knowing that in the paper on the Moody chart, Moody stated the accuracy is about ±5% for smooth pipes and ±10% for rough pipes. If more than one formula is applicable in the flow regime under consideration, the choice of formula may be influenced by one or more of the following:

- Required precision
- Speed of computation required
- Available computational technology:
  - —calculator (minimize keystrokes)
  - —spreadsheet (single-cell formula)
  - —programming/scripting language (subroutine).

Colebrook Equation

Compact Forms

The Colebrook equation is an implicit equation that combines experimental results of studies of turbulent flow in smooth and rough pipes. It was developed in 1939 by C. F. Colebrook . The 1937 paper by C. F. Colebrook and C. M. White is often erroneously cited as the source of the equation.

This is partly because Colebrook in a footnote (from his 1939 paper) acknowledges his debt to White for suggesting the mathematical method by which the smooth and rough pipe correlations could be combined. The equation is used to iteratively solve for the Darcy–Weisbach friction factor f. This equation is also known as the Colebrook–White equation.

For conduits that are flowing completely full of fluid at Reynolds numbers greater than 4000, it is defined as:

$$\frac{1}{\sqrt{f}} = -2\log_{10}\left(\frac{\varepsilon/D_{\mathrm{h}}}{3.7} + \frac{2.51}{\mathrm{Re}\sqrt{f}}\right)$$

or

$$\frac{1}{\sqrt{f}} = -2\log_{10}\left(\frac{\varepsilon}{14.8R_{\mathrm{h}}} + \frac{2.51}{\mathrm{Re}\sqrt{f}}\right)$$

where:

- f is the Darcy friction factor
- Roughness height, ε(m, ft)
- Hydraulic diameter, D_h (m, ft) — For fluid-filled, circular conduits, D_h = D = inside diameter
- Hydraulic radius, R_h (m, ft) — For fluid-filled, circular conduits, R_h = D/4 = (inside diameter)/4
- Re is the Reynolds number.

Solving Colebrook Equation

Due to the implicit nature of the Colebrook equation, determination of a friction factor requires some iteration or a numerical solving method.

Expanded forms

Additional, mathematically-equivalent forms of the Colebrook equation are:

$$\frac{1}{\sqrt{f}} = 1.7384... - 2\log_{10}\left(\frac{2\varepsilon}{D_\mathrm{h}} + \frac{18.574}{\mathrm{Re}\sqrt{f}}\right)$$

where:

1.7384... = 2 log (2 * 3.7) = 2 log (7.4)

18.574 = 2.51 * 3.7 * 2

and

$$\frac{1}{\sqrt{f}} = 1.1364... + 2\log_{10}(D_\mathrm{h}/\varepsilon) - 2\log_{10}\left(1 + \frac{9.287}{\mathrm{Re}(\varepsilon/D_\mathrm{h})\sqrt{f}}\right)$$

or

$$\frac{1}{\sqrt{f}} = 1.1364... - 2\log_{10}\left(\frac{\varepsilon}{D_\mathrm{h}} + \frac{9.287}{\mathrm{Re}\sqrt{f}}\right)$$

where:

1.1364... = 1.7384... - 2 log (2) = 2 log (7.4) - 2 log (2) = 2 log (3.7)

9.287 = 18.574 / 2 = 2.51 * 3.7.

The additional equivalent forms above assume that the constants 3.7 and 2.51 in the formula at the top of this section are exact. The constants are probably values which were rounded by Colebrook during his curve fitting; but they are effectively treated as exact when comparing (to several decimal places) results from explicit formulae (such as those found elsewhere in this article) to the friction factor computed via Colebrook's implicit equation.Equations similar to the additional forms above (with the constants rounded to fewer decimal places—or perhaps shifted slightly to minimize overall rounding errors) may be found in various references. It may be helpful to note that they are essentially the same equation.

Free surface flow

Another form of the Colebrook-White equation exists for free surfaces. Such a condition may exist in a pipe that is

flowing partially-full of fluid. For free surface flow:

$$\frac{1}{\sqrt{f}} = -2\log_{10}\left(\frac{\varepsilon}{12R_{\mathrm{h}}} + \frac{2.51}{\mathrm{Re}\sqrt{f}}\right).$$

Approximations of the Colebrook Equation

Haaland Equation

The *Haaland equation* is used to solve directly for the Darcy–Weisbach friction factor f for a full-flowing circular pipe. It is an approximation of the implicit Colebrook–White equation, but the discrepancy from experimental data is well within the accuracy of the data. It was developed by S. E. Haaland in 1983.

The Haaland equation is defined as:

$$\frac{1}{\sqrt{f}} = -1.8\log_{10}\left[\left(\frac{\varepsilon/D}{3.7}\right)^{1.11} + \frac{6.9}{\mathrm{Re}}\right]$$

where:

- f is the Darcy friction factor
- ε/D is the relative roughness
- Re is the Reynolds number.

Swamee–Jain Equation

The Swamee–Jain equation is used to solve directly for the Darcy–Weisbach friction factor f for a full-flowing circular pipe. It is an approximation of the implicit Colebrook–White equation.

where f is a function of:

- Roughness height, ε(m, ft)
- Pipe diameter, D(m, ft)
- Reynolds number, Re(unitless).

Serghides's Solution

Serghides's solution is used to solve directly for the Darcy–Weisbach friction factor f for a full-flowing circular pipe. It is an approximation of the implicit Colebrook–White equation. It was derived using Steffensen's method.

The solution involves calculating three intermediate values and then substituting those values into a final equation.

$$A = -2\log_{10}\left(\frac{\varepsilon/D}{3.7} + \frac{12}{\mathrm{Re}}\right)$$

$$B = -2\log_{10}\left(\frac{\varepsilon/D}{3.7} + \frac{2.51A}{\mathrm{Re}}\right)$$

$$C = -2\log_{10}\left(\frac{\varepsilon/D}{3.7} + \frac{2.51B}{\mathrm{Re}}\right)$$

$$f = \left(A - \frac{(B-A)^2}{C - 2B + A}\right)^{-2}$$

where f is a function of:

- Roughness height, ε(m, ft)
- Pipe diameter, D(m, ft)
- Reynolds number, Re(unitless).

The equation was found to match the Colebrook–White equation within 0.0023% for a test set with a 70-point matrix consisting of ten relative roughness values (in the range 0.00004 to 0.05) by seven Reynolds numbers (2500 to 10^8).

Goudar-Sonnad Equation

Goudar equation is the most accurate approximation to solve directly for the Darcy–Weisbach friction factor f for a full-flowing circular pipe. It is an approximation of the implicit Colebrook–White equation. Equation has the following form

$$a = \frac{2}{ln(10)}$$

$$b = \frac{\varepsilon/D}{3.7}$$

$$d = \frac{ln(10)Re}{5.02}$$

$$s = bd + ln(\mathrm{d})$$

$$q = s^{(s/(s+1))}$$

$$g = bd + ln\frac{d}{q}$$

$$z = ln\frac{q}{g}$$

$$D_{LA} = z\frac{g}{g+1}$$

$$D_{CFA} = D_{LA}(1 + \frac{z/2}{(g+1)^2 + (z/3)(2g-1)})$$

$$\frac{1}{\sqrt{f}} = a[ln(\frac{d}{q}) + D_{CFA}]$$

where f is a function of:

- Roughness height, ε(m, ft)
- Pipe diameter, D(m, ft)
- Reynolds number, Re(unitless).

Hazen–Williams Equation

The Hazen–Williams equation is an empirical formula which relates the flow of water in a pipe with the physical properties of the pipe and the pressure drop caused by friction. It is used in the design of water pipe systems such as fire sprinkler systems, water supply networks, and irrigation

systems. It is named after Allen Hazen and Gardner Stewart Williams. The Hazen–Williams equation has the advantage that the coefficient C is not a function of the Reynolds number, but it has the disadvantage that it is only valid for water. Also, it does not account for the temperature or viscosity of the water.

General Form

The general form of the equation relates the mean velocity of water in a pipe with the geometric properties of the pipe and slope of the energy line.

where:

- V is velocity
- k is a conversion factor for the unit system (k = 1.318 for US customary units, k = 0.849 for SI units)
- C is a roughness coefficient
- R is the hydraulic radius
- S is the slope of the energy line (head loss per length of pipe or h_f/L)

Typical C factors used in design, which take into account some increase in roughness as pipe ages are as follows:

Material	**C Factor low**	**C Factor high**	**Reference**
Asbestos-cement	140	140	-
Cast iron	100	140	-
Cement-Mortar Lined Ductile Iron Pipe	140	140	-
Concrete	100	140	-
Copper	130	140	-
Steel	90	110	-
Galvanized iron	120	120	-
Polyethylene	140	140	-
Polyvinyl chloride (PVC)	130	130	-
Fibre-reinforced plastic (FRP)	150	150	-

Pipe Equation

The general form can be specialized for full pipe flows. Taking the general form

$$V = k\,C\,R^{0.63}\,S^{0.54}$$

and exponentiating each side by 1 / 0.54 gives (rounding exponents to 2 decimals)

$$V^{1.85} = k^{1.85}\,C^{1.85}\,R^{1.17}\,S$$

Rearranging gives

$$S = \frac{V^{1.85}}{k^{1.85}\,C^{1.85}\,R^{1.17}}$$

The flow rate $Q = VA$, so

$$S = \frac{V^{1.85}A^{1.85}}{k^{1.85}\,C^{1.85}\,R^{1.17}\,A^{1.85}} = \frac{Q^{1.85}}{k^{1.85}\,C^{1.85}\,R^{1.17}\,A^{1.85}}$$

The hydraulic radius R (which is different from the geometric radius r) for a full pipe of geometric diameter d is $d/4$; the pipe's cross sectional area A is πd^2 / 4, so

$$S = \frac{4^{1.17}\,4^{1.85}\,Q^{1.85}}{\pi^{1.85}\,k^{1.85}\,C^{1.85}\,d^{1.17}\,d^{3.70}} = \frac{4^{3.02}\,Q^{1.85}}{\pi^{1.85}\,k^{1.85}\,C^{1.85}\,d^{4.87}}$$

$$= \frac{4^{3.02}}{\pi^{1.85}\,k^{1.85}}\,\frac{Q^{1.85}}{C^{1.85}\,d^{4.87}} = \frac{7.916}{k^{1.85}}\,\frac{Q^{1.85}}{C^{1.85}\,d^{4.87}}$$

Imperial Customary Units

When used to calculate the pressure drop using the Imperial customary customary units system, the equation is:

$$P_d = \frac{4.52}{C^{1.85}}\,\frac{Q^{1.85}}{d^{4.87}}$$

where:

P_d = pressure drop in pounds per square inch / foot

Q = flow in gallons per minute

$\underline{d}$ = inside pipe diameter (inch)

SI units

When used to calculate the pressure drop with the International System of Units, the equation becomes:

$$S = \frac{10.67}{C^{1.85}} \frac{Q^{1.85}}{d^{4.87}}$$

where:

- S = Head loss (in m of water) per m of pipeline
- Q = volumetric flow rate in m^3/s
- d = inside pipe diameter in m

Bernoulli's Principle

In fluid dynamics, Bernoulli's principle states that for an inviscid flow, an increase in the speed of the fluid occurs simultaneously with a decrease in pressure or a decrease in the fluid's potential energy. Bernoulli's principle is named after the Dutch-Swiss mathematician Daniel Bernoulli who published his principle in his book Hydrodynamica in 1738.

Bernoulli's principle can be applied to various types of fluid flow, resulting in what is loosely denoted as Bernoulli's equation. In fact, there are different forms of the Bernoulli equation for different types of flow. The simple form of Bernoulli's principle is valid for incompressible flows (e.g. most liquid flows) and also for compressible flows (*e.g.* gases) moving at low Mach numbers. More advanced forms may in some cases be applied to compressible flows at higher Mach numbers (see the derivations of the Bernoulli equation).

Bernoulli's principle can be derived from the principle of conservation of energy. This states that, in a steady flow, the sum of all forms of mechanical energy in a fluid along a streamline is the same at all points on that streamline. This requires that the

sum of kinetic energy and potential energy remain constant. If the fluid is flowing out of a reservoir the sum of all forms of energy is the same on all streamlines because in a reservoir the energy per unit mass (the sum of pressure and gravitational potential $\rho\, g\, h$) is the same everywhere.

Fluid particles are subject only to pressure and their own weight. If a fluid is flowing horizontally and along a section of a streamline, where the speed increases it can only be because the fluid on that section has moved from a region of higher pressure to a region of lower pressure; and if its speed decreases, it can only be because it has moved from a region of lower pressure to a region of higher pressure. Consequently, within a fluid flowing horizontally, the highest speed occurs where the pressure is lowest, and the lowest speed occurs where the pressure is highest.

Incompressible Flow Equation

In most flows of liquids, and of gases at low Mach number, the mass density of a fluid parcel can be considered to be constant, regardless of pressure variations in the flow. For this reason the fluid in such flows can be considered to be incompressible and these flows can be described as incompressible flow. Bernoulli performed his experiments on liquids and his equation in its original form is valid only for incompressible flow. A common form of Bernoulli's equation, valid at any arbitrary point along a streamline where gravity is constant, is:

$$\frac{v^2}{2} + gz + \frac{p}{\rho} = \text{constant}$$

where:

v is the fluid flow speed at a point on a streamline,

g is the acceleration due to gravity,

z is the elevation of the point above a reference plane, with the positive z-direction pointing upward—so in the direction opposite to the gravitational acceleration,

p is the pressure at the point, and

ρ is the density of the fluid at all points in the fluid.

For conservative force fields, Bernoulli's equation can be generalized as:

$$\frac{v^2}{2} + \Psi + \frac{p}{\rho} = \text{constant}$$

where Ψ is the force potential at the point considered on the streamline. *E.g.* for the Earth's gravity $\Psi = gz$. The following two assumptions must be met for this Bernoulli equation to apply:

- the fluid must be incompressible—even though pressure varies, the density must remain constant along a streamline;
- friction by viscous forces has to be negligible.

By multiplying with the fluid density ρ, equation (A) can be rewritten as:

$$\tfrac{1}{2}\rho v^2 + \rho g z + p = \text{constant}$$

or:

$$q + \rho g h = p_0 + \rho g z = \text{constant}$$

where:

$q = \tfrac{1}{2}\rho v^2$ is dynamic pressure,

$h = z + \dfrac{p}{\rho g}$ is the piezometric head or hydraulic head (the sum of the elevation z and the pressure head) and

$p_0 = p + q$ is the *total pressure* (the sum of the static pressure p and dynamic pressure q).

The constant in the Bernoulli equation can be normalised. A common approach is in terms of *total head* or *energy head* H:

$$H = z + \frac{p}{\rho g} + \frac{v^2}{2g} = h + \frac{v^2}{2g},$$

The above equations suggest there is a flow speed at which pressure is zero, and at even higher speeds the pressure is negative. Most often, gases and liquids are not capable of negative absolute pressure, or even zero pressure, so clearly Bernoulli's equation ceases to be valid before zero pressure is reached. In liquids—when the pressure becomes too low—cavitation occurs. The above equations use a linear relationship between flow speed squared and pressure. At higher flow speeds in gases, or for sound waves in liquid, the changes in mass density become significant so that the assumption of constant density is invalid.

Simplified Form

In many applications of Bernoulli's equation, the change in the $\rho\ g\ z$ term along the streamline is so small compared with the other terms it can be ignored. For example, in the case of aircraft in flight, the change in height z along a streamline is so small the $\rho\ g\ z$ term can be omitted. This allows the above equation to be presented in the following simplified form:

$$p + q = p_0$$

where p_0 is called total pressure, and q is dynamic pressure. Many authors refer to the pressure p as static pressure to distinguish it from total pressure p_0 and dynamic pressure q. In Aerodynamics, L.J. Clancy writes: "To distinguish it from the total and dynamic pressures, the actual pressure of the fluid, which is associated not with its motion but with its state,

is often referred to as the static pressure, but where the term pressure alone is used it refers to this static pressure."

The simplified form of Bernoulli's equation can be summarized in the following memorable word equation:

static pressure + dynamic pressure = total pressure

Every point in a steadily flowing fluid, regardless of the fluid speed at that point, has its own unique static pressure p and dynamic pressure q. Their sum $p + q$ is defined to be the total pressure p_0. The significance of Bernoulli's principle can now be summarized as *total pressure is constant along a streamline*.

If the fluid flow is irrotational, the total pressure on every streamline is the same and Bernoulli's principle can be summarized as *total pressure is constant everywhere in the fluid flow*. It is reasonable to assume that irrotational flow exists in any situation where a large body of fluid is flowing past a solid body. Examples are aircraft in flight, and ships moving in open bodies of water. However, it is important to remember that Bernoulli's principle does not apply in the boundary layer or in fluid flow through long pipes.

If the fluid flow at some point along a stream line is brought to rest, this point is called a stagnation point, and at this point the total pressure is equal to the stagnation pressure.

Applicability of incompressible Flow Equation to Flow of Gases

Bernoulli's equation is sometimes valid for the flow of gases: provided that there is no transfer of kinetic or potential energy from the gas flow to the compression or expansion of the gas. If both the gas pressure and volume change simultaneously, then work will be done on or by the gas. In this case, Bernoulli's equation—in its incompressible flow form—can not be assumed to be valid. However if the gas process is entirely isobaric, or isochoric, then no work is done

on or by the gas, (so the simple energy balance is not upset). According to the gas law, an isobaric or isochoric process is ordinarily the only way to ensure constant density in a gas. Also the gas density will be proportional to the ratio of pressure and absolute temperature, however this ratio will vary upon compression or expansion, no matter what non-zero quantity of heat is added or removed. The only exception is if the net heat transfer is zero, as in a complete thermodynamic cycle, or in an individual isentropic (frictionless adiabatic) process, and even then this reversible process must be reversed, to restore the gas to the original pressure and specific volume, and thus density. Only then is the original, unmodified Bernoulli equation applicable. In this case the equation can be used if the flow speed of the gas is sufficiently below the speed of sound, such that the variation in density of the gas (due to this effect) along each streamline can be ignored. Adiabatic flow at less than Mach 0.3 is generally considered to be slow enough.

Unsteady Potential Flow

The Bernoulli equation for unsteady potential flow is used in the theory of ocean surface waves and acoustics.

For an irrotational flow, the flow velocity can be described as the gradient $\Delta\phi$ of a velocity potential ϕ. In that case, and for a constant density ρ, the momentum equations of the Euler equations can be integrated to:

$$\frac{\partial\varphi}{\partial t} + \tfrac{1}{2}v^2 + \frac{p}{\rho} + gz = f(t),$$

which is a Bernoulli equation valid also for unsteady—or time dependent—flows. Here $\partial\phi/\partial t$ denotes the partial derivative of the velocity potential ϕ with respect to time t, and $v = |\Delta\phi|$ is the flow speed. The function f(t) depends only on time and not on position in the fluid. As a result, the Bernoulli equation at some moment t does not only apply along a certain streamline,

but in the whole fluid domain. This is also true for the special case of a steady irrotational flow, in which case f is a constant.

Further *f*(t) can be made equal to zero by incorporating it into the velocity potential using the transformation

$$\Phi = \varphi - \int_{t_0}^{t} f(\tau)\,\mathrm{d}\tau, \text{ resulting in } \frac{\partial \Phi}{\partial t} + \tfrac{1}{2}v^2 + \frac{p}{\rho} + gz = 0.$$

Note that the relation of the potential to the flow velocity is unaffected by this transformation: $\Delta\phi = \Delta\phi$.

The Bernoulli equation for unsteady potential flow also appears to play a central role in Luke's variational principle, a variational description of free-surface flows using the Lagrangian (not to be confused with Lagrangian coordinates).

Compressible Flow Equation

Bernoulli developed his principle from his observations on liquids, and his equation is applicable only to incompressible fluids, and compressible fluids at very low speeds (perhaps up to 1/3 of the sound speed in the fluid). It is possible to use the fundamental principles of physics to develop similar equations applicable to compressible fluids. There are numerous equations, each tailored for a particular application, but all are analogous to Bernoulli's equation and all rely on nothing more than the fundamental principles of physics such as Newton's laws of motion or the first law of thermodynamics.

Compressible Flow in Fluid Dynamics

For a compressible fluid, with a barotropic equation of state, and under the action of conservative forces,

$$\frac{v^2}{2} + \int_{p1}^{p} \frac{d\tilde{p}}{\rho(\tilde{p})} + \Psi = \text{constant} \quad \text{(constant along a streamline)}$$

where:

p is the pressure

ρ is the density

v is the flow speed

Ψ is the potential associated with the conservative force field, often the gravitational potential

In engineering situations, elevations are generally small compared to the size of the Earth, and the time scales of fluid flow are small enough to consider the equation of state as adiabatic. In this case, the above equation becomes

$$\frac{v^2}{2} + gz + \left(\frac{\gamma}{\gamma - 1}\right)\frac{p}{\rho} = \text{constant (constant along a streamline)}$$

where, in addition to the terms listed above:

γ is the ratio of the specific heats of the fluid

g is the acceleration due to gravity

z is the elevation of the point above a reference plane

In many applications of compressible flow, changes in elevation are negligible compared to the other terms, so the term gz can be omitted. A very useful form of the equation is then:

$$\frac{v^2}{2} + \left(\frac{\gamma}{\gamma - 1}\right)\frac{p}{\rho} = \left(\frac{\gamma}{\gamma - 1}\right)\frac{p_0}{\rho_0}$$

where:

p_0 is the total pressure

ρ_0 is the total density

Compressible Flow in Thermodynamics

Another useful form of the equation, suitable for use in thermodynamics, is:

$$\frac{v^2}{2} + \Psi + w = \text{constant}.$$

Here w is the enthalpy per unit mass, which is also often written as h (not to be confused with "head" or "height").

$$w = \epsilon + \frac{p}{\rho}$$

Note that where ε is the thermodynamic energy per unit mass, also known as the specific internal energy or "sie."

The constant on the right hand side is often called the Bernoulli constant and denoted b. For steady inviscid adiabatic flow with no additional sources or sinks of energy, b is constant along any given streamline. More generally, when b may vary along streamlines, it still proves a useful parameter, related to the "head" of the fluid (see below).

When the change in Ψ can be ignored, a very useful form of this equation is:

$$\frac{v^2}{2} + w = w_0$$

where w_0 is total enthalpy. For a calorically perfect gas such as an ideal gas, the enthalpy is directly proportional to the temperature, and this leads to the concept of the total (or stagnation) temperature.

When shock waves are present, in a reference frame in which the shock is stationary and the flow is steady, many of the parameters in the Bernoulli equation suffer abrupt changes in passing through the shock. The Bernoulli parameter itself, however, remains unaffected. An exception to this rule is radiative shocks, which violate the assumptions leading to the Bernoulli equation, namely the lack of additional sinks or sources of energy.

Derivations of Bernoulli Equation

Bernoulli Equation for Incompressible Fluids

The Bernoulli equation for incompressible fluids can be derived by integrating the Euler equations, or applying the law of conservation of energy in two sections along a streamline, ignoring viscosity, compressibility, and thermal effects. The simplest derivation is to first ignore gravity and consider constrictions and expansions in pipes that are otherwise straight, as seen in Venturi effect. Let the x axis be directed down the axis of the pipe. The equation of motion for a parcel of fluid, having a length $\mathrm{d}x$, mass density ρ, mass $m = \rho A\, \mathrm{d}x$ and flow velocity $\mathrm{v} = \mathrm{d}x / \mathrm{d}t$, moving along the axis of the horizontal pipe, with cross-sectional area A and change in pressure dp is

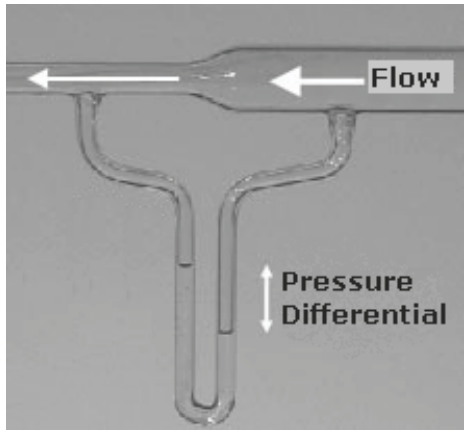


Fig 10—Showing the Venturi effect

$$m\frac{\mathrm{d}v}{\mathrm{d}t} = F$$

$$\rho A\mathrm{d}x\frac{\mathrm{d}v}{\mathrm{d}t} = -A\mathrm{d}p$$

$$\rho\frac{\mathrm{d}v}{\mathrm{d}t} = -\frac{\mathrm{d}p}{\mathrm{d}x}$$

In steady flow, $v = v(x) = v(x(t))$, so v itself is not directly a function of time t. Only through the cross-sectional position $x(t)$ it is that v depends from t:

$$\frac{\mathrm{d}v}{\mathrm{d}t} = \frac{\mathrm{d}v}{\mathrm{d}x}\frac{\mathrm{d}x}{\mathrm{d}t} = \frac{\mathrm{d}v}{\mathrm{d}x}v = \frac{d}{\mathrm{d}x}\left(\frac{v^2}{2}\right).$$

With density ρ constant, the equation of motion can be written as

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\rho\frac{v^2}{2} + p\right) = 0$$

or

$$\frac{v^2}{2} + \frac{p}{\rho} = C$$

where C is a constant, sometimes referred to as the Bernoulli constant. It is not a universal constant, but rather a constant of a particular fluid system. The deduction is: where the speed is large, pressure is low and vice versa. In the above derivation, no external work-energy principle is invoked. Rather, Bernoulli's principle was inherently derived by a simple manipulation of the momentum equation.

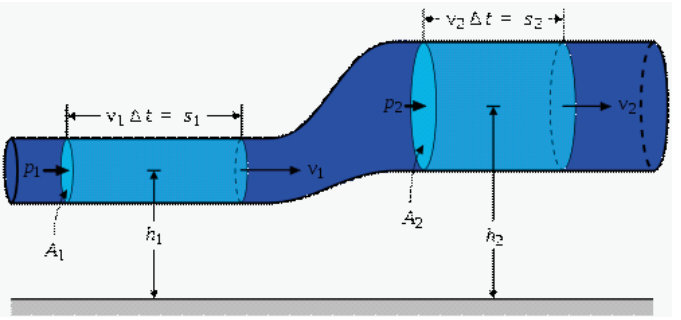


Fig 11—A streamtube of fluid moving to the right. Indicated are pressure, elevation, flow speed, distance (s), and cross-sectional area. Note that in this figure elevation is denoted as h, contrary to the text where it is given by z.

Another way to derive Bernoulli's principle for an incompressible flow is by applying conservation of energy. In the form of the work-energy theorem, stating that

the change in the kinetic energy E_{kin} of the system equals the net work $\mathbb{W}$ done on the system;

$$W = \Delta E_{\text{kin}}.$$

Therefore,

the work done by the forces in the fluid = increase in kinetic energy.

The system consists of the volume of fluid, initially between the cross-sections A_1 and A_2. In the time interval Δt fluid elements initially at the inflow cross-section A1 move over a distance $s_1 = v_1 \Delta t$, while at the outflow cross-section the fluid moves away from cross-section A_2 over a distance $s_2 = v_2 \Delta t$. The displaced fluid volumes at the inflow and outflow are respectively $A_1 s_1$ and $A_2 s_2$. The associated displaced fluid masses are—when ρ is the fluid's mass density—equal to density times volume, so $\rho A_1 s_1$ and $\rho A_2 s_2$. By mass conservation, these two masses displaced in the time interval Δt have to be equal, and this displaced mass is denoted by Δm:

$$\rho A_1 s_1 = \rho A_1 v_1 \Delta t = \Delta m,$$
$$\rho A_2 s_2 = \rho A_2 v_2 \Delta t = \Delta m.$$

The work done by the forces consists of two parts:

- The work done by the pressure acting on the area's A_1 and A_2

$$W_{\text{pressure}} = F_{1,\text{pressure}}\, s_1 - F_{2,\text{pressure}}\, s_2 = p_1 A_1 s_1 - p_2 A_2 s_2 = \Delta m \frac{p_1}{\rho} - \Delta m \frac{p}{\rho}$$

- The work done by gravity: the gravitational potential energy in the volume $A_1 s_1$ is lost, and at the outflow in the volume $A_2 s_2$ is gained. So, the change in gravitational potential energy $\Delta E_{\text{pot,gravity}}$ in the time interval Δt is

$$\Delta E_{\text{pot,gravity}} = \Delta m\, g z_2 - \Delta m\, g z_1.$$

Now, the work by the force of gravity is opposite to the

change in potential energy, $W_{\text{gravity}} = -\Delta E_{\text{pot,gravity}}$: while the force of gravity is in the negative z-direction, the work—gravity force times change in elevation—will be negative for a positive elevation change $\Delta z = z_2 - z_1$, while the corresponding potential energy change is positive. So:

$$W_{\text{gravity}} = -\Delta E_{\text{pot,gravity}} = \Delta m\, g z_1 - \Delta m\, g z_2.$$

And the total work done in this time interval Δt is

$$W = W_{\text{pressure}} + W_{\text{gravity}}.$$

The increase in kinetic energy is

$$\Delta E_{\text{kin}} = \frac{1}{2}\Delta m\, v_2^2 - \frac{1}{2}\Delta m\, v_1^2.$$

Putting these together, the work-kinetic energy theorem

$$\Delta m \frac{p_1}{\rho} - \Delta m \frac{p_2}{\rho} + \Delta m\, g z_1 - \Delta m\, g z_2 = \frac{1}{2}\Delta m\, v_2^2 - \frac{1}{2}\Delta m\, v_1^2$$

or

$$\frac{1}{2}\Delta m\, v_1^2 + \Delta m\, g z_1 + \Delta m \frac{p_1}{\rho} = \frac{1}{2}\Delta m\, v_2^2 + \Delta m\, g z_2 + \Delta m \frac{p_2}{\rho}.$$

After dividing by the mass $\Delta m = \rho\, A_1\, v_1\, \Delta t = \rho\, A_2\, v_2\, \Delta_t$ the result is:

or, as stated in the first paragraph:

$$\frac{v^2}{2} + g z + \frac{p}{\rho} = C$$

Further division by g produces the following equation. Note that each term can be described in the length dimension (such as meters). This is the head equation derived from Bernoulli's principle:

$$\frac{v^2}{2g} + z + \frac{p}{\rho g} = C$$

The middle term, z, represents the potential energy of

the fluid due to its elevation with respect to a reference plane. Now, z is called the elevation head and given the designation $z_{\text{elevation}}$.

A free falling mass from an elevation $z > 0$ (in a vacuum) will reach a speed $v = \sqrt{2gz}$, when arriving at elevation $z = 0$. Or when we rearrange it as a head: $h_v = \frac{v^2}{2g}$ The term $v^2 / (2g)$ is called the *velocity head*, expressed as a length measurement. It represents the internal energy of the fluid due to its motion.

The hydrostatic pressure p is defined as

$p = p_0 - \rho g z$, with p_0 some reference pressure, or when we rearrange it as a head: $\psi = \frac{p}{\rho g}$

The term $p / (\rho g)$ is also called the *pressure head*, expressed as a length measurement. It represents the internal energy of the fluid due to the pressure exerted on the container.

When we combine the head due to the flow speed and the head due to static pressure with the elevation above a reference plane, we obtain a simple relationship useful for incompressible fluids using the velocity head, elevation head, and pressure head.

$$h_v + z_{\text{elevation}} + \psi = C$$

If we were to multiply Eqn. 1 by the density of the fluid, we would get an equation with three pressure terms:

$$\frac{\rho v^2}{2} + \rho g z + p = C$$

We note that the pressure of the system is constant in this form of the Bernoulli Equation. If the static pressure of the system (the far right term) increases, and if the pressure due to elevation (the middle term) is constant, then we know that the dynamic pressure (the left term) must have decreased. In other words, if the speed of a fluid decreases and it is not due to an elevation difference, we know it must be due to an increase in the static pressure that is resisting the flow.

All three equations are merely simplified versions of an energy balance on a system.

Bernoulli Equation for Compressible Fluids

The derivation for compressible fluids is similar. Again, the derivation depends upon (1) conservation of mass, and (2) conservation of energy. Conservation of mass implies that in the above figure, in the interval of time Δt, the amount of mass passing through the boundary defined by the area A_1 is equal to the amount of mass passing outwards through the boundary defined by the area A_2:

$$0 = \Delta M_1 - \Delta M_2 = \rho_1 A_1 v_1 \, \Delta t - \rho_2 A_2 v_2 \, \Delta t$$

Conservation of energy is applied in a similar manner: It is assumed that the change in energy of the volume of the streamtube bounded by A1 and A2 is due entirely to energy entering or leaving through one or the other of these two boundaries. Clearly, in a more complicated situation such as a fluid flow coupled with radiation, such conditions are not met. Nevertheless, assuming this to be the case and assuming the flow is steady so that the net change in the energy is zero,

$$0 = \Delta E_1 - \Delta E_2$$

where ΔE_1 and ΔE_2 are the energy entering through A_1 and leaving through A_2, respectively.

The energy entering through A1 is the sum of the kinetic energy entering, the energy entering in the form of potential gravitational energy of the fluid, the fluid thermodynamic energy entering, and the energy entering in the form of mechanical $p\,dV$ work:

$$\Delta E_1 = \left[\frac{1}{2}\rho_1 v_1^2 + \Psi_1 \rho_1 + \epsilon_1 \rho_1 + p_1\right] A_1 v_1 \, \Delta t$$

where Ψ = gz is a force potential due to the Earth's gravity, g is acceleration due to gravity, and z is elevation above a reference plane.

A similar expression for ΔE_2 may easily be constructed. So now setting $0 = \Delta E_1 - \Delta E_2$:

$$0 = \left[\frac{1}{2}\rho_1 v_1^2 + \Psi_1\rho_1 + \epsilon_1\rho_1 + p_1\right] A_1 v_1\,\Delta t - \left[\frac{1}{2}\rho_2 v_2^2 + \Psi_2\rho_2 + \epsilon_2\rho_2 + p_2\right] A_2 v_2\,\Delta t$$

which can be rewritten as:

$$0 = \left[\frac{1}{2}v_1^2 + \Psi_1 + \epsilon_1 + \frac{p_1}{\rho_1}\right]\rho_1 A_1 v_1\,\Delta t - \left[\frac{1}{2}v_2^2 + \Psi_2 + \epsilon_2 + \frac{p_2}{\rho_2}\right]\rho_2 A_2 v_2\,\Delta t$$

Now, using the previously-obtained result from conservation of mass, this may be simplified to obtain

$$\frac{1}{2}v^2 + \Psi + \epsilon + \frac{p}{\rho} = \text{constant} \equiv b$$

which is the Bernoulli equation for compressible flow.

Chapter: Two

Biology at Nanointerface

Molecules and Macromolecules

Small organic molecules are the basic stuff of life. Scientists call them *monomers*. Mono means one. These monomers actually make up building blocks for bigger things. You can think of them as one brick that makes up a brick wall. When monomers (small molecules) are joined together, they form larger molecules called polymers. Poly means many. Think of the *polymers* as that brick wall. And when polymers are joined together, they form "giant" molecules called *macromolecules*. Macro means big. You can think of these macromolecules as the building that is made up of those brick walls. So, you need the bricks (monomers) to form the walls (polymers) which when put together actually make up the building (macromolecules). Macromolecules are what this Instruction is mostly going to be about.

But first we'd probably better review a little basic chemistry. That's because even though this lesson is called Cell Biology, we're really talking *biochemistry* here Don't let the term freak you out. It is just the chemistry of living things (like cells). So here's a little dictionary to remind you about some of the words that we'll be using in this Instruction. Don't worry, you don't have to memorize all of these terms. Use them for reference later if you have questions while reading this lesson.

You will actually run across words like compound, molecule and bond.

Words from Chemistry

atom:-The smallest component of an element that has all the properties of that element. An atom is made up of protons, neutrons and electrons. The electrons play an important role in chemical bonding.

element: -A substance composed of atoms that are alike. For example, oxygen is made up of two atoms of oxygen (O). This means that the formula for oxygen is O_2.

compound:- A substance composed of atoms that are different. Water is a good example, since water is made up of two atoms of hydrogen (H) plus one of oxygen (O). The chemical formula for water is H_2O.

molecule:- The smallest part of a substance that can exist on its own without losing its chemical properties. Molecule of an element consists of two or more of the same atoms bonded together, like the two atoms that make up oxygen.

The molecule of a compound consists of two or more different atoms bonded together (like water, H_2O).

bond:- When atoms, elements or compounds join together they bond by sharing their electrons. The electrons of certain atoms only bond with the electrons of certain other atoms.

formula:-A written representation or structural model that tells you or shows you which elements go into what compound and how. There are many different kinds of formulae (formulae is the plural of formula) but the two most useful for our purpose are:

- the *molecular formula*, in which the chemical compounds are written like a sentence, and
- the Kekule model in which they are drawn and look something like a graph.

We'll use a simple example to show you what we mean. Methane gas is made up of one atom of carbon and four atoms of hydrogen.

So the molecular (written) formula would be CH_4.

Now let's get back to macromolecules. Just refer back to the chemistry terms if you get confused about the meaning of one of them.

Macromolecules

All macromolecules (polymers) are produced from a very small group of about 50 monomers (molecules). These monomers can be put together to form many different substances. But the process by which they're put together is always the same. And so is the process by which they're broken apart to release energy.

Polymers are put together by an anabolic (building) process called dehydration synthesis. Sounds scary right? Well, actually when you translate the words, dehydration is just removal of water, and synthesis just means to build. In dehydration synthesis, two molecules are chemically bonded through the use of enzymes and the removal of water. So, you build something and take water away in the process.

Polymers are broken apart by a catabolic (destruction) process called hydrolysis. Another potentially scary term right? Wrong, all you need to do is break the word down. Hydro refers to water and lyse is to break. You are breaking something apart and adding water. Let's see how this all works with something we've just been learning about—carbohydrates (sugars).

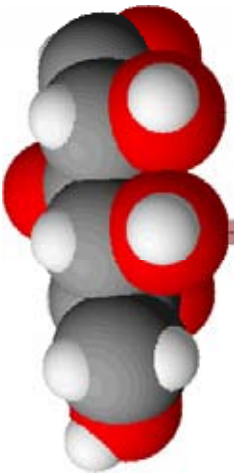
Carbohydrates

Carbohydrates (sugars) and their polymers are the main source of food for all living things. Chemists call these sugars saccharides.

This brings us to something you've probably already noticed. Different kinds of scientists have different words for the same thing. They have their reasons, but it can be pretty confusing.

Carbohydrates (sugars or saccharides) are macromolecules made up of carbon, oxygen and hydrogen. The basic formula for carbohydrates is CH_2O. You have your simple sugars, which are just made up of single (mono) sugars, when you put a couple of these things together, you form two (di) sugars. If you put many (poly) saccharides together you get an even bigger saccharide.

- monosaccharides (single sugars)—CH_2O (glucose, fructose and galactose)
- disaccharides (double sugars)—$C_{12}H_{22}0_{11}$ (maltose, sucrose and lactose)
- polysaccharides (multiple sugars; poly means many)—$C_6H_{10}O_5n$ (starch and fiber; found in grain products, fruits and vegetables)



Monosaccharides and disaccharides (simple carbohydrates) are the most important source of nutrition for cells and bodies.

Polysaccharides (complex carbohydrates) do the work of building and storage.

Cellulose, which is the most abundant organic compound on earth, is a basic building material for plants. No, we didn't say cellulite—but cellulite is made up of macromolecules, too.

Here is a diagram of a carbohydrate:

Another important macromolecule is fat.

Fat

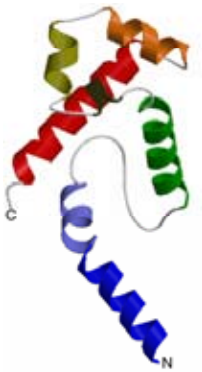
Fats are large molecules composed of 2 types of monomers—glycerol (an alcohol containing carbons) and 3 fatty acid molecules. Fatty acid contains oxygen, hydrogen and carbon. The bond connecting the glycerol and the fatty acids in the fat molecule is called an ester bond.

There are two types of fatty acid: *saturated* and *unsaturated*. The saturated fatty acids do not contain a double bond between their carbon atoms, while the unsaturated fatty acids do contain one or more double bonds between carbon atoms. You can read about them on any food label. If you see "unsaturated fats" on the label, it is referring to mostly plant fats. If you read "saturated fats" on the label, it is a reference to mostly animal fats. Which is better? Consider the fact that diets that are high in saturated fats have been linked to cardiovascular disease. These types of fats contribute to something called, artherosclerosis. It is pronounced, art heroes clear o sis. It is a big word that describes fatty deposits that build up on the insides of blood vessels. That means it will slow down blood flow. Also, beware of "hydrogentated fats" on your labels. What you see to describe this in the grocery store is the term "trans fats." These fats are called hydrogenated because you take unsaturated fats and add hydrogen to make them saturated fats. These contribute to high cholesterol. There are a couple of kinds of cholesterol, one is good, one is bad, we will learn about those later. Trans fats contribute to a rise in bad cholesterol.

Protein

You should recall from an earlier instruction that proteins are the most diverse of all molecules. There are seven major classes of proteins. You don't have to learn all of them here, but you do need to know that they make up things like hair, silk, antibodies, hormones and of course enzymes. You should remember learning all about enzymes from an earlier lesson. Like carbohydrates and lipids, proteins are made up of carbon (C), hydrogen (H) and oxygen (O). In addition to the C, the H and the O, they also contain nitrogen (N). All of these elements are arranged to form amino acids. Amino acids are the building blocks of proteins. Proteins are formed when these amino acids are linked in a chain. So if one amino acid (peptide) binds to another by dehydration synthesis, a dipeptide is formed. Do you remember, dehydration synthesis puts things together, while taking away water?

All amino acids (also called peptides) have the same basic structure—a central carbon (C) atom with a hydrogen (H) attached to it. Also bound to this central carbon is something called an "R" group. It is a strange name, but the "R" group refers to a bunch of different chemicals. You don't have to learn all of these chemical groups, just know that there are different ones, and they are all referred to as R groups. When the R group changes, so does the amino acid. So look at the protein diagram and notice this central carbon with the hydrogen attached, also hanging off of the carbon is this R group. This central carbon (with the H and the "R" group) also connects to an amino group (NH_2), and an acid group. It



The nucleic acid molecule

makes much more sense if you look at the diagram. Here is a diagram of the basic monomer of protein. It is also an example of a monomer that would be attached to other monomers by a peptide (amino acid) bond.

Nucleotides

There are two types of nucleic acids. Deoxyribonucleic Acid, more commonly known as DNA, and ribonucleic acid, more commonly known as RNA. You should recall from an earlier lesson what these nucleic acids do. They provide all of the instruction for proteins. Organisms inherit this DNA from their parents. Your curly hair or blue eyes come directly from the genes that your parents or their parents passed on to you. The genes are found inside the DNA. As you recall from an earlier lesson, the DNA cannot pass its information directly, it needs RNA. This all happens pretty much in the same way you snail mail (not email) a letter. Your letter, the DNA gets picked up by the mailman (RNA) and eventually gets delivered to its final destination. The DNA is important, but it couldn't get where it needs to be without the RNA. So, now that you remember how DNA transmits it message, let's look at what makes it up.

The smallest units (monomers) of nucleic acids are called *nucleotides*. How would you build a nucleotide? Good question, each one has three parts.

- Part one: Take one *five-carbon sugar*
- Part two: One *phosphate group*
- Part Three: One *nitrogenous base*

Put the five-carbon sugar in the middle. Attach a phosphate group to one end and a nitrogenous base to the other end and you have a nucleotide!

But, you notice that you have two types of nucleic acids. Is the recipe the same for both? Both are composed of

nucleotides. These nucleotides are very similar, with a few minor changes. Lets look at the differences.

First, consider the names; Deoxyribonucleic Acid, Ribonucleic Acid.

For *part one*, the deoxyribose is the five-carbon sugar you would use to form DNA's nucleotide. Ribose is the five-carbon sugar you would use to build RNA.

For *part two*, the phosphate groups are the same for both forms of nucleic acids. There are no differences here.

For *part three*, the nitrogenous bases are very similar for both DNA and RNA. There are four kinds of nitrogenous bases that can be found in DNA.

DNA's nitrogenous bases are:

Adenine=A

Thymine=T

Cytosine=C

Guanine=G

These nitrogenous bases pair up to help form the famous double helix you may have heard about with DNA. The A bonds to T and the C bonds to G. The nitrogenous bases of RNA are very similar except instead of Thymine, RNA contains Uracil (U).

RNA's nitrogenous bases are:

Adenine=A

Uracil=U

Cytosine=C

Guanine=G

Remember that these nucleotides are the monomers that make up the nucleic acids. To put a couple of these nucleotides together you have to (you guessed it) take out some water to

form a bond, dehydration synthesis. The phosphate group from one nucleotide will bind to the sugar in the next. This forms a sugar-phosphate backbone. The nitrogenous bases protrude into the center. Proteins are polymers made up of amino acids. They are the most complex and important group of molecules because of all the different functions they perform to support life. Every cell that makes up a plant or an animal requires proteins for its structure and function. In this classroom activity, students will learn about the sources of proteins and their uses in the food industry. In part I, students will precipitate casein from milk using an acid (this is the method used to make cottage cheese). In Part 2, students will coagulate casein from milk using an enzyme (this is the method used to make cheese). And in Part 3, students will coagulate soy protein from soymilk, using magnesium sulfate (this is the method used to make tofu). Note that the products of these experiments are not to be eaten. Molecules are involved in everything -- including in how you smell and taste. Odor and food molecules activate membrane receptors in your nose and mouth. Each substance we smell or taste has a unique chemical signature although most substances are made up of a number of different molecules. Humans have hundreds of kinds of odor membrane receptors and perhaps 50 to 100 kinds of taste receptors. Although we typically describe only five categories of taste -- salty, sour, sweet, bitter, and umami (the taste of monosodium glutamate and similar molecules)—each category probably has more than one type of receptor. In this experiment, students learn how to investigate the senses of taste and smell and also find out how to plan and carry out their own experiments. In the Class Experiment, students find that the ability to identify a flavor depends on the sense of smell as well as the sense of taste. They also learn basic facts about food molecules, sensory receptors, nerve connections and brain centers.

Chapter: Three

Biology of Nano Signalling

Cells—receive and act on signals Signal brings about response

Types of signals:

Autocrine

Paracrine

Endocrine

Lots of signals but just a few evolutionarily conserved mechanisms to detect signals and transduce them into change in cell

I. Ligand-gated Ion Channel

Nicotinic Acetylcholine receptor

*O*pens in response to neurotransmitter acetylcholine and to nicotine

Found in neurons and muscle fibers

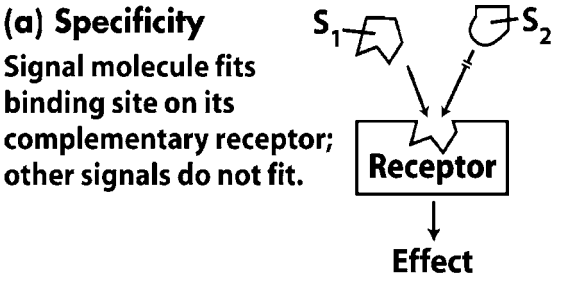
II. Receptor Enzymes

Insulin Receptor

Ligand-binding domain on extracellular surface of plasma membrane. Enzyme active site on cytosolic side

III. G protein-coupled Receptors and Second Messengers

β-Adrenergic Receptor



Weak interactions
Receptor cell-specific
High affinity of receptors for signal
Cooperativity

Fig. 1—Showing the Bio signaling

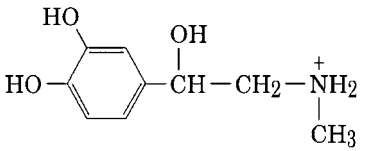
a.k.a. Adrenaline

Regulates metabolism in muscle, liver and fat

Breakdown of glycogen and fat

G protein-coupled Receptors and Second Messengers

Epinephrine



Epinephrine

IV. Steroid Receptors

Act in nucleus to alter gene expression Steroid hormones (estrogen, progesterone, cortisol, etc.) hydrophobic Receptors (proteins) and HREs (hormone response elements in DNA)

Steroid Receptors

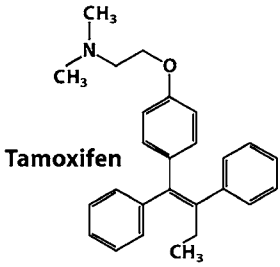
Receptor for estrogen

Breast cancer—some types need estrogen present for tumor growth

Tamoxifen = antagonist of estrogen

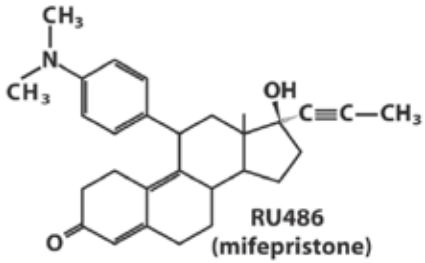
Tamoxifen competes with estrogen for binding to receptor

Tamoxifen has no effect on gene expression like estrogen does



RU486 = antagonist of progesterone Competes with prog for binding to receptor

Prog needed for proper implantation of fertilized ovum in uterus



Oncogenes, Tumor Suppressor Genes, Programmed Cell Death

Tumors—result of uncontrolled cell division - biosignaling gone BAD!

Oncogenes—a cancer-causing gene, any of several mutant genes that cause cells to exhibit rapid, uncontrolled proliferation Discovered in tumor-causing viruses Very similar to normal genes in the body called proto-oncogenes (growth regulating genes)

Tumor Suppressor Genes—encode proteins that normally restrain cell division, mutation in one or more can lead to tumor growth p53 - mutated in 90% skin cancers, 50% all other cancers Rb - mutated in retinoblastoma.

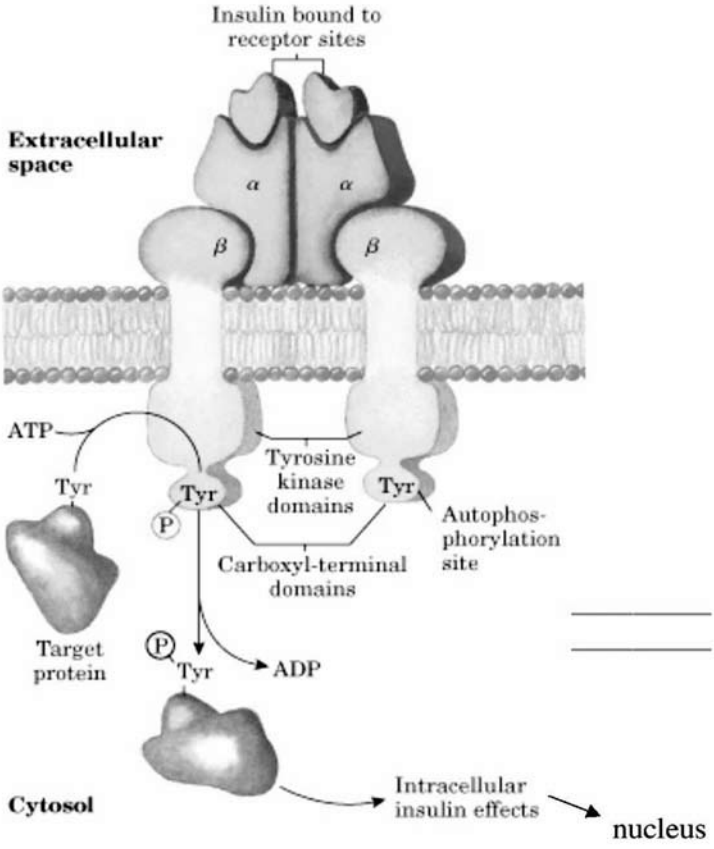


Fig. 2—Showing the Signaling in the endocrine hormones

II. Receptor Enzymes
Insulin Receptor

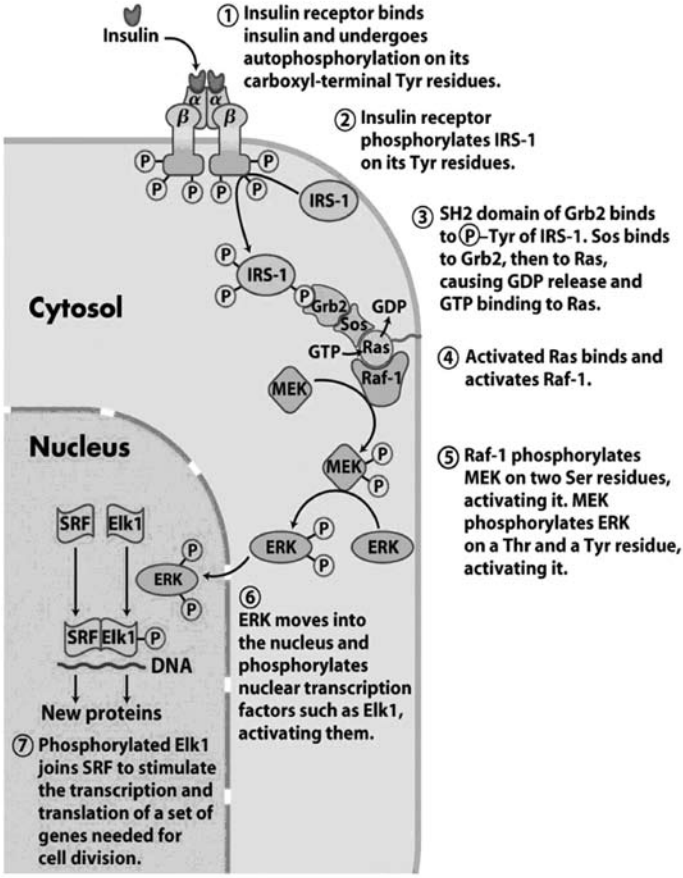
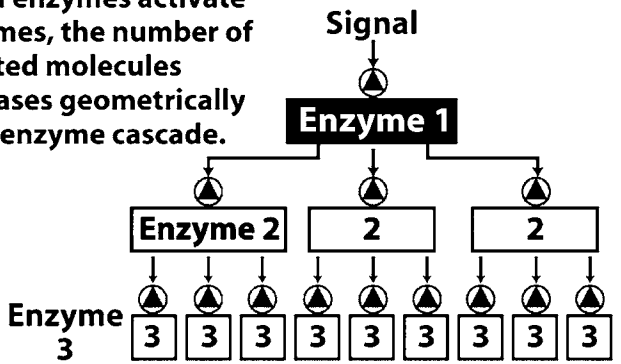


Fig. 3—Showing the Signalling activies in the insulin receptor

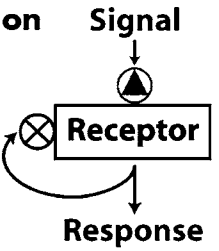
(b) Amplification

When enzymes activate enzymes, the number of affected molecules increases geometrically in an enzyme cascade.



(c) Desensitization/Adaptation

Receptor activation triggers a feedback circuit that shuts off the receptor or removes it from the cell surface.



(d) Integration

When two signals have opposite effects on a metabolic characteristic such as the concentration of a second messenger X, or the membrane potential V_m, the regulatory outcome results from the integrated input from both receptors.

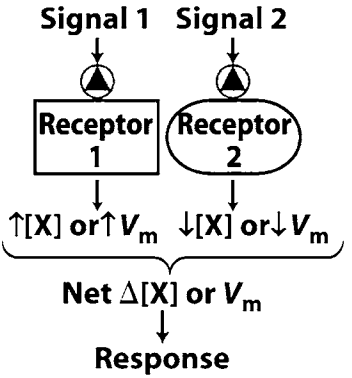


Fig. 4—The Process of Biosignaling

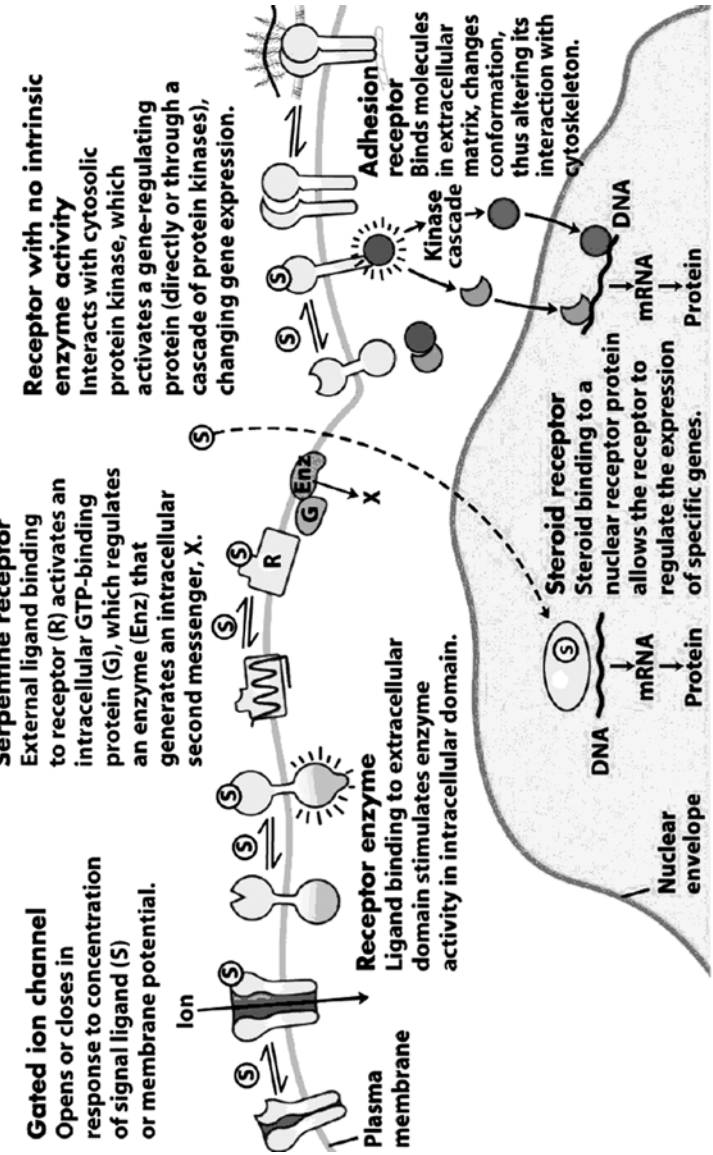


Fig. 5—Showing the Signalling between the nuclic acids

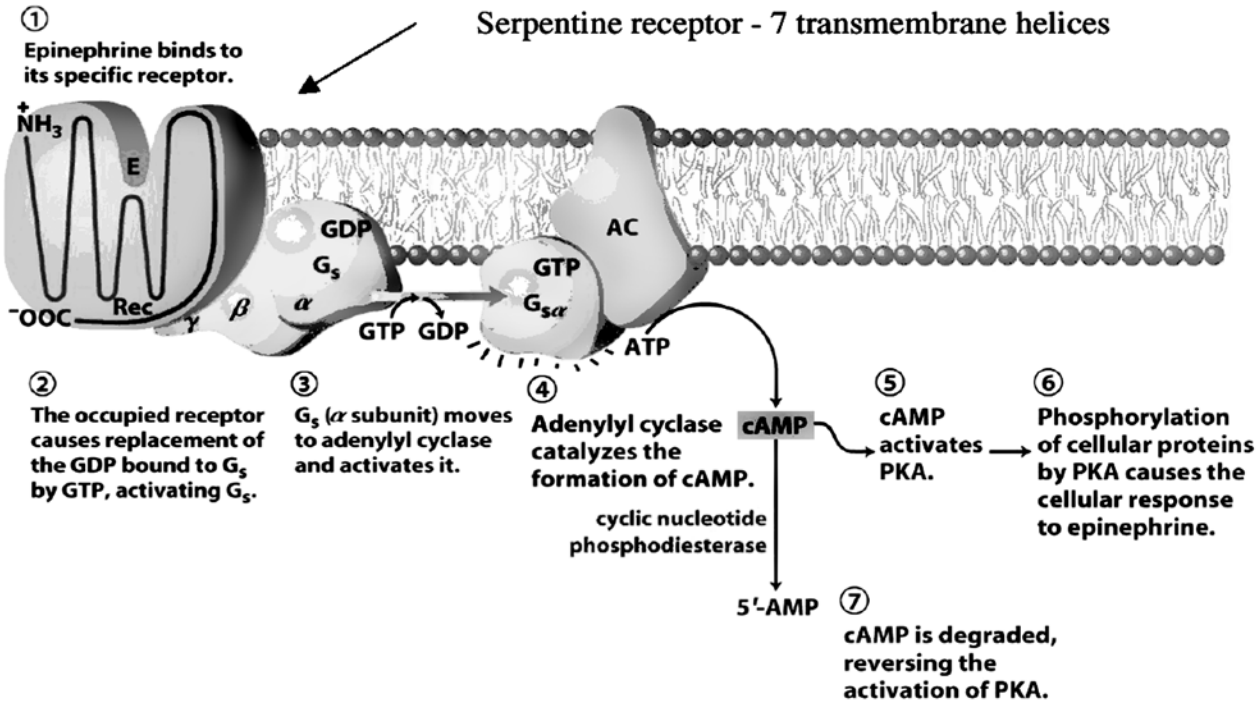


Fig. 6—Showing the Signaling Process in the Serpentine receptors

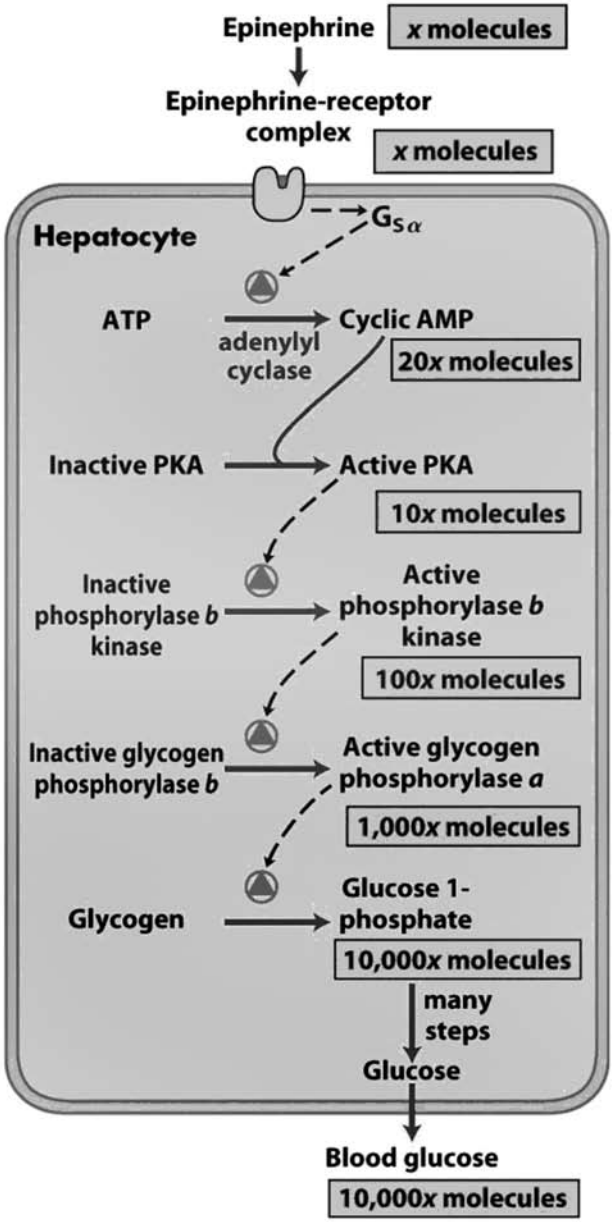
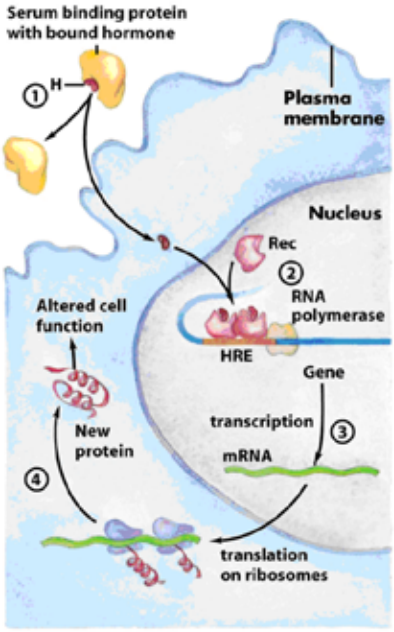


Fig. 7—The Epinephrine Signaling



1. Hormone (H), carried to the target tissue on serum binding proteins, diffuses across the plasma membrane and binds to its specific receptor protein (Rec) in the nucleus.
2. Hormone binding changes the conformation of Rec; it forms homo- or heterodimers with other hormone-receptor complexes and binds to specific regulatory regions called hormone response elements (HREs) in the DNA adjacent to specific genes.
3. Binding regulates transcription of the adjacent gene(s), increasing or decreasing the rate of mRNA formation.
4. Altered levels of the hormone-regulated gene product produce the cellular response to the hormone.

Fig. 8—Showing the Nucleic Acid Signaling

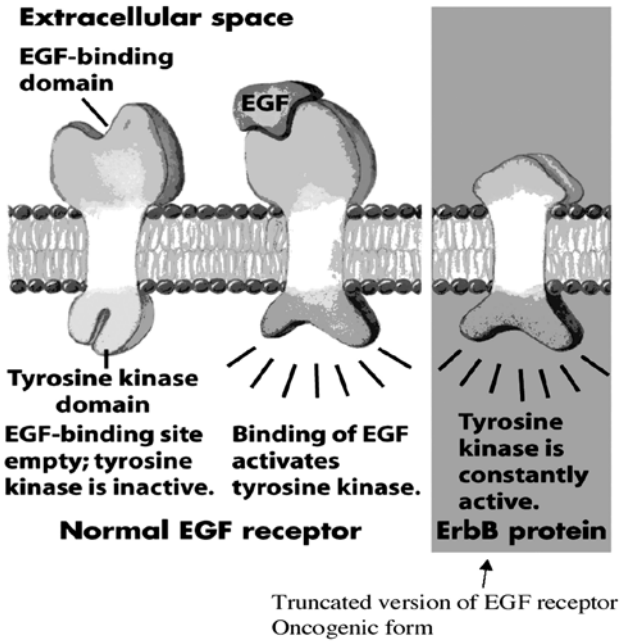


Fig. 9—Showing EGP signalling

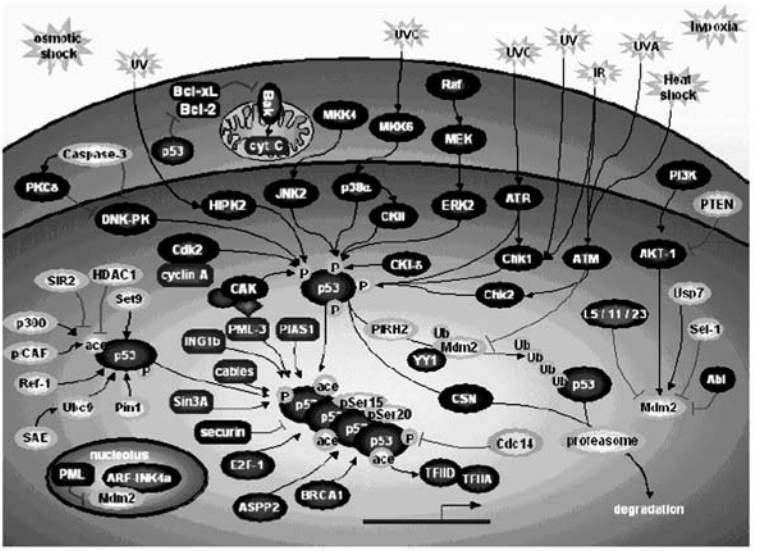


Fig. 10—The siganling in the Oncogens

Oncogenes, Tumor Suppressor Genes, Programmed Cell Death Tumor Progression

Programmed Cell Death (Apoptosis)—cell brings about its own death and lysis, signaled from outside or programmed in its genes, by systematically degrading its own macromolecules

When?

Development of embryo (fingers)

Anti-self antibodies present

Menstruation Stressed cells (virus-infected to prevent infection, heat, UV light)

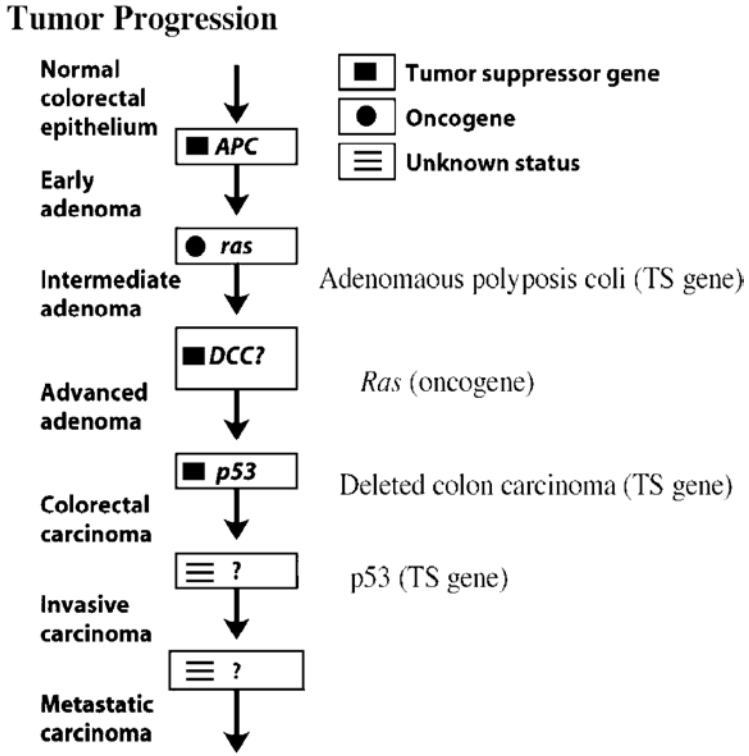


Fig. 11—Showing the flow chart of tumor in progress

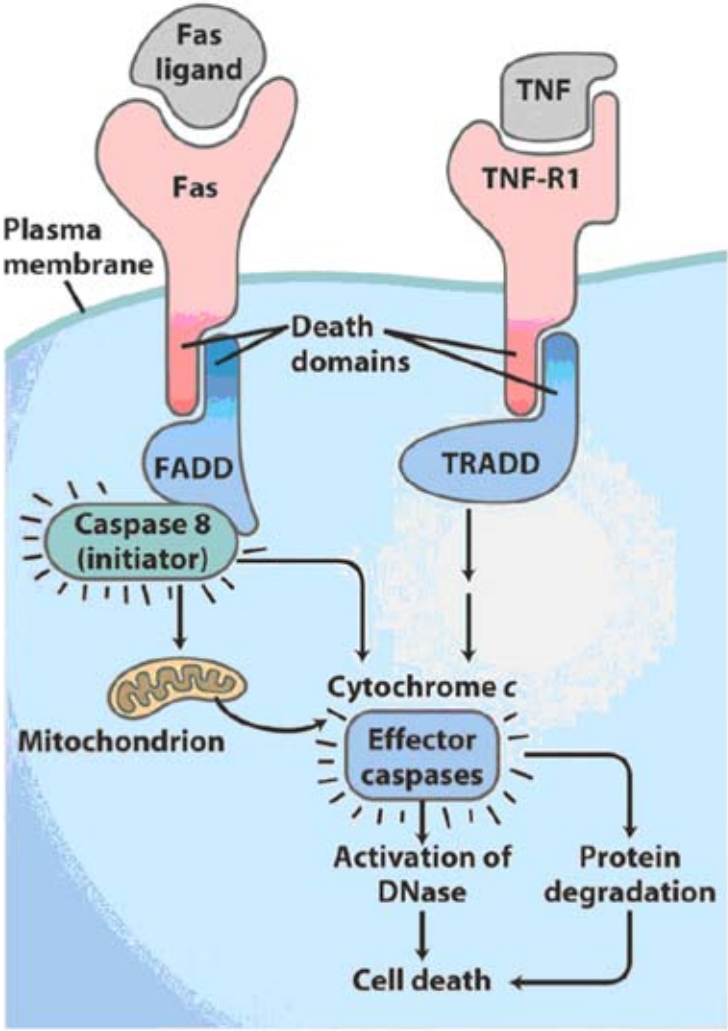


Fig. 12—Showing the Mutation in tumers

Mutation to any of these Proteins can Lead to Cancers

Cell signaling is part of a complex system of communication that governs basic cellular activities and coordinates cell actions. The ability of cells to perceive and correctly respond to their microenvironment is the basis of development, tissue repair, and immunity as well as normal tissue homeostasis. Errors in cellular information processing are responsible for diseases such as cancer, autoimmunity, and diabetes. By understanding cell signaling, diseases may be treated effectively and, theoretically, artificial tissues may be created.Traditional work in biology has focused on studying individual parts of cell signaling pathways. Systems biology research helps us to understand the underlying structure of cell signaling networks and how changes in these networks may affect the transmission and flow of information. Such networks are complex systems in their organization and may exhibit a number of emergent properties including bistability and ultrasensitivity. Analysis of cell signaling networks requires a combination of experimental and theoretical approaches including the development and analysis of simulations and modelling.

Unicellular and Multicellular Organism Cell Signaling

Figure 13. Example of signaling between bacteria. Salmonella enteritidis uses acyl-homoserine lactone for Quorum sensing (see: Inter-Bacterial Communication)

Cell signaling has been most extensively studied in the context of human diseases and signaling between cells of a single organism. However, cell signaling may also occur between the cells of two different organisms. In many mammals, early embryo cells exchange signals with cells of the uterus. In the human gastrointestinal tract, bacteria exchange signals with each other and with human epithelial and immune system

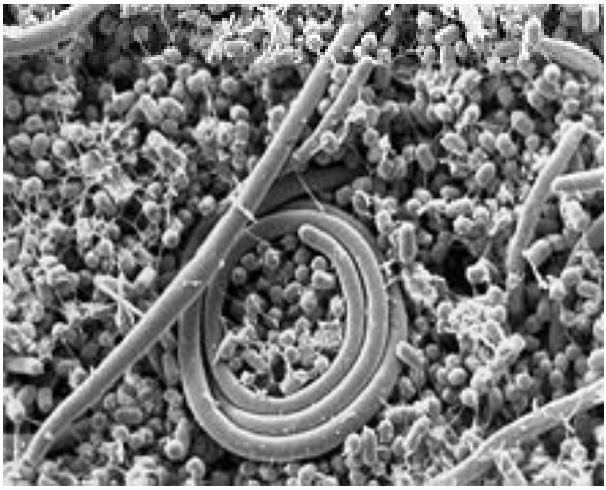


Fig. 13—Example of signaling between bacteria.

cells. For the yeast Saccharomyces cerevisiae during mating, some cells send a peptide signal (mating factor pheromones) into their environment. The mating factor peptide may bind to a cell surface receptor on other yeast cells and induce them to prepare for mating.

Types of Signals

Cells communicate with each other via direct contact (juxtacrine signaling), over short distances (paracrine signaling), or over large distances and/or scales (endocrine signaling). Some cell-to-cell communication requires direct cell-cell contact. Some cells can form gap junctions that connect

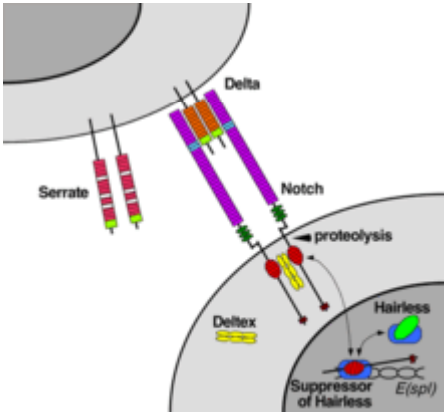


Fig. 14—Notch-mediated juxtacrine signal between adjacent cells.

their cytoplasm to the cytoplasm of adjacent cells. In cardiac muscle, gap junctions between adjacent cells allows for action potential propagation from the cardiac pacemaker region of the heart to spread and coordinately cause contraction of the heart.

The Notch Signalling

The Notch signaling mechanism is an example of juxtacrine signalling (also known as contact-dependent signaling) in which two adjacent cells must make physical contact in order to communicate. This requirement for direct contact allows for very precise control of cell differentiation during embryonic development. In the worm Caenorhabditis elegans, two cells of the developing gonad each have an equal chance of terminally differentiating or becoming a uterine precursor cell that continues to divide. The choice of which cell continues to divide is controlled by competition of cell surface signals. One cell will happen to produce more of a cell surface protein that activates the Notch receptor on the adjacent cell. This activates a feedback loop or system that reduces Notch expression in the cell that will differentiate and that increases Notch on the surface of the cell that continues as a stem cell. Many cell signals are carried by molecules that are released by one cell and move to make contact with another cell. Endocrine signals are called hormones. Hormones are produced by endocrine cells and they travel through the blood to reach all parts of the body. Specificity of signaling can be controlled if only some cells can respond to a particular hormone. Paracrine signals such as retinoic acid target only cells in the vicinity of the emitting cell. Neurotransmitters represent another example of a paracrine signal. Some signaling molecules can function as both a hormone and a neurotransmitter. For example, epinephrine and norepinephrine can function as hormones when released from the adrenal gland and are transported

to the heart by way of the blood stream. Norepinephrine can also be produced by neurons to function as a neurotransmitter within the brain. Estrogen can be released by the ovary and function as a hormone or act locally via paracrine or autocrine signaling. Active species of oxygen and nitric oxide can also act as cellular messengers. This process is dubbed redox signaling.

Receptors for Cell Signals

Cells receive information from their environment through a class of proteins known as receptors. Notch is a cell surface protein that functions as a receptor. Animals have a small set of genes that code for signaling proteins that interact specifically with Notch receptors and stimulate a response in cells that express Notch on their surface. Molecules that activate (or, in some cases, inhibit) receptors can be classified as hormones, neurotransmitters, cytokines, growth factors but all of these are called receptor ligands. The details of ligand-receptor interactions are fundamental to cell signaling. Notch acts as a receptor for ligands that are expressed on adjacent cells. While many receptors are cell surface proteins, some are found inside cells. For example, oestrogen is a hydrophobic molecule that can pass through the lipid bilayer of cell surface membranes. Oestrogen receptors inside cells of the uterus can be activated by oestrogen that comes from the ovaries, enters the target cells, and binds to oestrogen receptors. A number of transmembrane receptors for molecules that include peptide hormones and of intracellular receptors for steroid hormones exist, giving to a cell the ability to respond to a great number of hormonal and pharmacological stimuli. In diseases, often, proteins that interact with receptors are aberrantly activated, resulting in constitutively activated downstream signals. For several types of intercellular signaling molecules that are unable to permeate the hydrophobic cell membrane due to

their hydrophilic nature, the target receptor is expressed on the membrane. When such signaling molecule activates its receptor, the signal is carried into the cell usually by means of a second messenger such as cAMP.

Signaling Pathways

In some cases, receptor activation caused by ligand binding to a receptor is directly coupled to the cell's response to the ligand. For example, the neurotransmitter GABA can activate a cell surface receptor that is part of an ion channel. GABA binding to a GABA A receptor on a neuron opens a chloride-selective ion channel that is part of the receptor. GABA A receptor activation allows negatively-charged chloride ions to move into the neuron, which inhibits the ability of the neuron to produce action potentials. However, for many cell surface receptors, ligand-receptor interactions are not directly linked to the cell's response. The activated receptor must first interact with other proteins inside the cell before the ultimate physiological effect of the ligand on the cell's behavior is produced. Often, the behavior of a chain of several interacting cell proteins is altered following receptor activation. The entire set of cell changes induced by receptor activation is called a signal transduction mechanism or pathway.

In the case of Notch-mediated signaling, the signal transduction mechanism can be relatively simple. As shown in left, activation of Notch can cause the Notch protein to be altered by a protease. Part of the Notch protein is released from the cell surface membrane and can act to change the pattern of gene transcription in the cell nucleus. This causes the responding cell to make different proteins, resulting in an altered pattern of cell behavior. Cell signaling research involves studying the spatial and temporal dynamics of both receptors and the components of signaling pathways that are activated by receptors in various cell types.

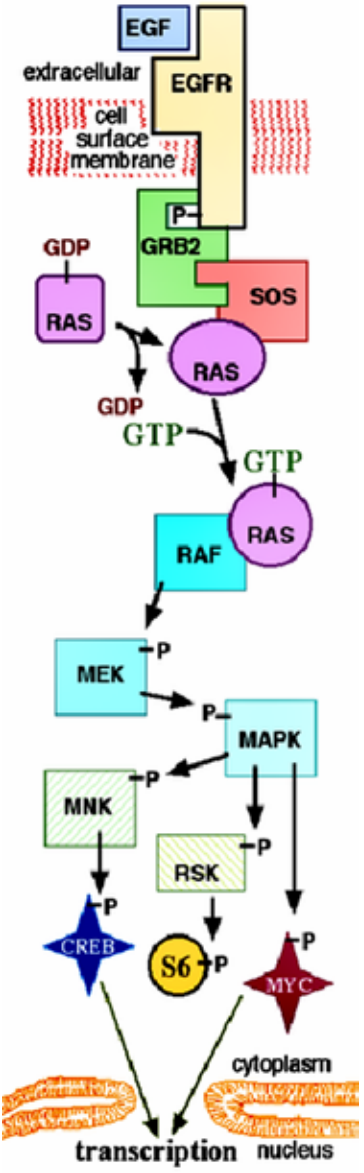


Fig. 15A—Overview of signal transduction pathways.

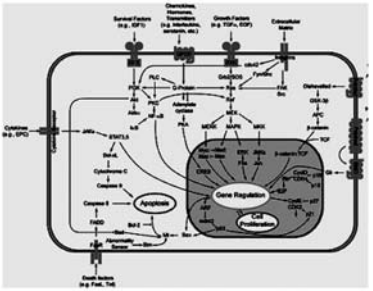


Fig. 15—Diagram showing key components of a signal transduction pathway. See the MAPK/ERK pathway article for details.

A more complex signal transduction pathway is shown in figure. This pathway involves changes of protein-protein interactions inside the cell, induced by an external signal. Many growth factors bind to receptors at the cell surface and stimulate cells to progress through the cell cycle and divide. Several of these receptors are kinases that start to phosphorylate themselves and other proteins when binding to a ligand. This phosphorylation can generate a binding site for a different protein and thus induce protein-protein interaction. The ligand (called epidermal growth factor (EGF)) binds to the receptor (called EGFR). This activates the

receptor to phosphorylate itself. The phosphorylated receptor binds to an adaptor protein (GRB2), which couples the signal to further downstream signaling processes. For example, one of the signal transduction pathways that are activated is called the mitogen-activated protein kinase (MAPK) pathway. The signal transduction component labeled as "MAPK" in the pathway was originally called "ERK," so the pathway is called the MAPK/ERK pathway. The MAPK protein is an enzyme, a protein kinase that can attach phosphate to target proteins such as the transcription factor MYC and, thus, alter gene transcription and, ultimately, cell cycle progression. Many cellular proteins are activated downstream of the growth factor receptors (such as EGFR) that initiate this signal transduction pathway.

Some signaling transduction pathways respond differently depending on the amount of signaling received by the cell. For instance, the hedgehog protein activates different genes, depending on the amount of hedgehog protein present.

Complex multi-component signal transduction pathways provide opportunities for feedback, signal amplification, and interactions inside one cell between multiple signals and signaling pathways.

Classification of Intercellular Communication

Within endocrinology (the study of intercellular signalling in animals) and the endocrine system, intercellular signalling is subdivided into the following classifications:

- *Intracrine* signals are produced within the target cell.
- *Autocrine* signals target the cell itself. Sometimes autocrine cells can target cells close by if they are the same type of cell as the emitting cell. An example of this is immune cells.
- *Juxtacrine* signals target adjacent (touching) cells.

These signals are transmitted along cell membranes via protein or lipid components integral to the membrane and are capable of affecting either the emitting cell or cells immediately adjacent.

- *Paracrine* signals target cells in the vicinity of the emitting cell. Neurotransmitters represent an example.
- *Endocrine* signals target distant cells. Endocrine cells produce hormones that travel through the blood to reach all parts of the body.

Other Bio Signaling Strategies

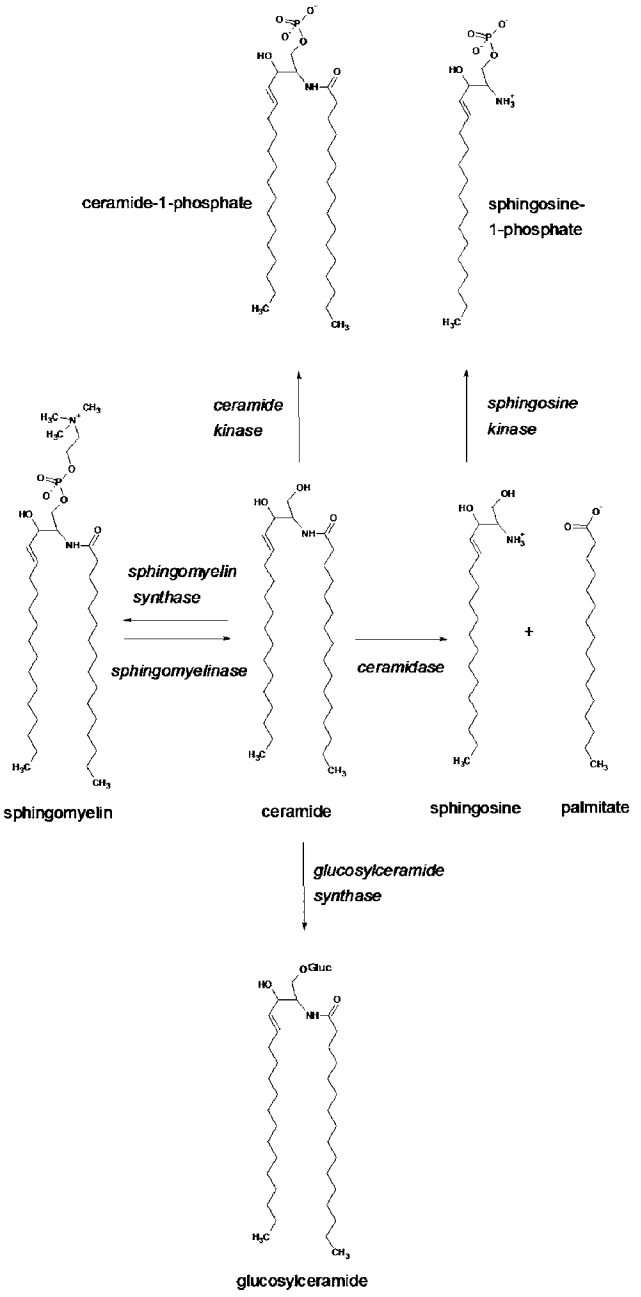
Lipid signaling, broadly defined, refers to any biological signaling event involving a lipid messenger that binds a protein target, such as a receptor, kinase or phosphatase, which in turn mediate the effects of these lipids on specific cellular responses. Lipid signaling is thought to be qualitatively different from other classical signaling paradigms (such as monoamine neurotransmission) because lipids can freely diffuse through membranes (see osmosis.) One consequence of this is that lipid messengers cannot be stored in vesicles prior to release and so are often biosynthesized "on demand" at their intended site of action. As such, many lipid signaling molecules cannot circulate freely in solution but, rather, exist bound to special carrier proteins in serum.

Sphingolipid Second Messengers

Sphingolipid second messengers. Ceramide is at the metabolic hub, leading to the formation of other sphingolipids.

Ceramide

Ceramide (Cer) can be generated by the breakdown of sphingomyelin (SM) by sphingomyelinases (SMases), which


ceramide-1-phosphate
sphingosine-
1-phosphate
ceramide
kinase
sphingosine
kinase
sphingomyelin
synthase
sphingomyelinase
ceramidase
+
sphingomyelin
ceramide
sphingosine
palmitate
glucosylceramide
synthase
glucosylceramide

are enzymes that hydrolyze the phosphocholine group from the sphingosine backbone. Alternatively, this sphingosine-derived lipid (sphingolipid) can be synthesized from scratch (de novo) by the enzymes serine palmitoyl transferase (SPT) and ceramide synthase in organelles such as the endoplasmic reticulum (ER) and possibly, in the mitochondria-associated membranes (MAMs) and the perinuclear membranes. Being located in the metabolic hub, ceramide leads to the formation of other sphingolipids, with the C1 hydroxyl (-OH) group as the major site of modification. A sugar can be attached to ceramide (glycosylation) through the action of the enzymes, glucosyl or galactosyl ceramide synthases. Ceramide can also be broken down by enzymes called ceramidases, leading to the formation of sphingosine, Moreover, a phosphate group can be attached to ceramide (phosphorylation) by the enzyme, ceramide kinase.

It is also possible to regenerate sphingomyelin from ceramide by accepting a phosphocholine headgroup from phosphatidylcholine (PC) by the action of an enzyme called sphingomyelin synthase.[5] The latter process results in the formation of diacylglycerol (DAG) from PC.Ceramide contains two hydrophobic ("water-fearing") chains and a neutral headgroup. Consequently, it has limited solubility in water and is restricted within the organelle where it was formed. Also, because of its hydrophobic nature, ceramide readily flip-flops across membranes as supported by studies in membrane models and membranes from red blood cells (erythrocytes).[6] However, ceramide can possibly interact with other lipids to form bigger regions called microdomains which restrict its flip-flopping abilities. This could have immense effects on the signaling functions of ceramide because it is known that ceramide generated by acidic SMase enzymes in the outer leaflet of an organelle membrane may have different roles compared to ceramide that is formed in the inner leaflet by the

action of neutral SMase enzymes. Ceramide mediates many cell-stress responses, including the regulation of programmed cell death (apoptosis) and cell aging (senescence). Numerous research works have focused interest on defining the direct protein targets of action of ceramide. These include enzymes called ceramide-activated Ser-Thr phosphatases (CAPPs), such as protein phosphatase 1 and 2A (PP1 and PP2A), which were found to interact with ceramide in studies done in a controlled environment outside of a living organism (in vitro). On the other hand, studies in cells have shown that ceramide-inducing agents such as tumor necrosis factor-alpha α (TNFα) and palmitate induce the ceramide-dependent removal of a phosphate group (dephosphorylation) of the retinoblastoma gene product RB and the enzymes, protein kinases B (AKT protein family) and C α (PKB and PKCα). Moreover, there is also sufficient evidence which implicates ceramide to the activation of the kinase suppressor of Ras (KSR), PKCζ, and cathepsin D. Interestingly, cathepsin D has been proposed as the main target for ceramide formed in organelles called lysosomes, making lysosomal acidic SMase enzymes one of the key players in the mitochondrial pathway of apoptosis. Ceramide was also shown to activate PKCζ, implicating it to the inhibition of AKT, regulation of the voltage difference between the interior and exterior of the cell (membrane potential) and signaling functions that favor apoptosis. Chemotherapeutic agents such as daunorubicin and etoposide enhance the de novo synthesis of ceramide in studies done on mammalian cells.

The same results were found for certain inducers of apoptosis particularly stimulators of receptors in a class of lymphocytes (a type of white blood cell) called B-cells.[20] Regulation of the de novo synthesis of ceramide by palmitate may have a key role in diabetes and the metabolic syndrome. Experimental evidence shows that there is

substantial increase of ceramide levels upon adding palmitate. Ceramide accumulation activates PP2A and the subsequent dephosphorylation and inactivation of AKT, a crucial mediator in metabolic control and insulin signaling. This results in a substantial decrease in insulin responsiveness (*i.e.* to glucose) and in the death of insulin-producing cells in the pancreas called islets of Langerhans. Inhibition of ceramide synthesis in mice via drug treatments or gene-knockout techniques prevented insulin resistance induced by fatty acids, glucocorticoids or obesity. An increase in in vitro activity of acid SMase has been observed after applying multiple stress stimuli such as ultraviolet (UV) and ionizing radiation, binding of death receptors and chemotherapeutic agents such as platinum, histone deacetylase inhibitors and paclitaxel. In some studies, SMase activation results to its transport to the plasma membrane and the simultaneous formation of ceramide. Ceramide transfer protein (CERT) transports ceramide from ER to the Golgi for the synthesis of SM. CERT is known to bind phosphatidylinositol phosphates, hinting its potential regulation via phosphorylation, a step of the ceramide metabolism that can be enzymatically regulated by protein kinases and phosphatases, and by inositol lipid metabolic pathways. Up to date, there are at least 26 distinct enzymes with varied subcellular localizations, that act on ceramide as either a substrate or product. Regulation of ceramide levels can therefore be performed by one of these enzymes in distinct organelles by particular mechanisms at various times.

Sphingosine

Sphingosine (Sph) is formed by the action of ceramidase (CDase) enzymes on ceramide in the lysosome. Sph can also be formed in the extracellular (outer leaflet) side of the plasma membrane by the action of neutral CDase enzyme. Sph then is either recycled back to ceramide or phosphorylated by one

of the sphingosine kinase enzymes, SK1 and SK2. The product sphingosine-1-phosphate (S1P) can be dephosphorylated in the ER to regenerate sphingosine by certain S1P phosphatase enzymes within cells, where the salvaged Sph is recycled to ceramide. Sphingosine is a single-chain lipid (usually 18 carbons in length), rendering it to have sufficient solubility in water. This explains its ability to move between membranes and to flip-flop across a membrane. Estimates conducted at physiological pH show that approximately 70% of sphingosine remains in membranes while the remaining 30% is water-soluble. Sph that is formed has sufficient solubility in the liquid found inside cells (cytosol). Thus, Sph may come out of the lysosome and move to the ER without the need for transport via proteins or membrane-enclosed sacs called vesicles. However, its positive charge favors partitioning in lysosomes. It is proposed that the role of SK1 located near or in the lysosome is to 'trap' Sph via phosphorylation. It is important to note that since sphingosine exerts surfactant activity, it is one of the sphingolipids found at lowest cellular levels.

The low levels of Sph and their increase in response to stimulation of cells, primarily by activation of ceramidase by growth-inducing proteins such as platelet-derived growth factor and insulin-like growth factor, is consistent with its function as a second messenger. It was found that immediate hydrolysis of only 3 to 10% of newly-generated ceramide may double the levels of Sph. Treatment of HL60 cells (a type of leukemia cell line) by a plant-derived organic compound called phorbol ester increased Sph levels three-fold, whereby the cells differentiated into white blood cells called macrophages. Treatment of the same cells by exogenous Sph caused apoptosis. A specific protein kinase phosphorylates 14-3-3, otherwise known as sphingosine-dependent protein kinase 1 (SDK1), only in the presence of Sph. Sph is also known to interact

with protein targets such as the protein kinase H homologue (PKH) and the yeast protein kinase (YPK). These targets in turn mediate the effects of Sph and its related sphingoid bases, with known roles in regulating the actin cytoskeleton, endocytosis, the cell cycle and apoptosis. It is important to note however that the second messenger function of Sph is not yet established unambiguously.

Sphingosine-1-Phosphate

Sphingosine-1-phosphate (S1P), like Sph, is composed of a single hydrophobic chain and has sufficient solubility to move between membranes. S1P is formed by phosphorylation of sphingosine by sphingosine kinase (SK). The phosphate group of the product can be detached (dephosphorylated) to regenerate sphingosine via S1P phosphatase enzymes or S1P can be broken down by S1P lyase enzymes to ethanolamine phosphate and hexadecenal. Similar to Sph, its second messenger function is not yet clear. However, there is substantial evidence that implicates S1P to cell survival, cell migration, and inflammation. Certain growth-inducing proteins such as platelet-derived growth factor (PDGF), insulin-like growth factor (IGF) and vascular endothelial growth factor (VEGF) promote the formation of SK enzymes, leading to increased levels of S1P. Other factors that induce SK include cellular communication molecules called cytokines, such as tumor necrosis factor α (TNFα) and interleukin-1 (IL-1), hypoxia or lack of oxygen supply in cells, oxidized low-density lipoproteins (oxLDL) and several immune complexes. S1P is probably formed at the inner leaflet of the plasma membrane in response to TNFα and other receptor activity-altering compounds called agonists. S1P, being present in low nanomolar concentrations in the cell, has to interact with high-affinity receptors that are capable of sensing their low levels. So far, the only identified receptors for S1P are the high-affinity G protein-coupled

receptors (GPCRs), also known as S1P receptors (S1PRs). S1P is required to reach the extracellular side (outer leaflet) of the plasma membrane to interact with S1PRs and launch typical GPCR signaling pathways. However, the zwitterionic headgroup of S1P makes it unlikely to flip-flop spontaneously. To overcome this difficulty, the ATP-binding cassette (ABC) transporter C1 (ABCC1) serves as the "exit door" for S1P. On the other hand, the cystic fibrosis transmembrane regulator (CFTR) serves as the means of entry for S1P into the cell. In contrast to its low intracellular concentration, S1P is found in high nanomolar concentrations in serum where it is bound to albumin and lipoproteins. Inside the cell, S1P can induce calcium release independent of the S1PRs—the mechanism of which remains unknown. To date, the intracellular molecular targets for S1P are still unidentified.

The SK1-S1P pathway has been extensively studied in relation to cytokine action, with multiple functions connected to effects of TNFα and IL-1 favoring inflammation. Studies show that knockdown of key enzymes such as S1P lyase and S1P phosphatase increased prostaglandin production, parallel to increase of S1P levels. This strongly suggests that S1P is the mediator of SK1 action and not subsequent compounds. Research done on endothelial and smooth muscle cells is consistent to the hypothesis that S1P has a crucial role in regulating endothelial cell growth, and movement. Recent work on a sphingosine analogue, FTY270, demonstrates its ability to act as a potent compound that alters the activity of S1P receptors (agonist). FTY270 was further verified in clinical tests to have roles in immune modulation, such as that on multiple sclerosis. This highlights the importance of S1P in the regulation of lymphocyte function and immunity. Most of the studies on S1P are used to further understand diseases such as cancer, arthritis and inflammation, diabetes, immune function and neurodegenerative disorders.

Glucosylceramide

Glucosylceramides (GluCer) are the most widely distributed glycosphingolipids in cells serving as precursors for the formation of over 200 known glycosphingolipids. GluCer is formed by the glycosylation of ceramide in an organelle called Golgi via enzymes called glucosylceramide synthase (GCS) or by the breakdown of complex glycosphingolipids (GSLs) through the action of specific hydrolase enzymes. In turn, certain β-glucosidases hydrolyse these lipids to regenerate ceramide. GluCer appears to be synthesized in the inner leaflet of the Golgi. Studies show that GluCer has to flip to the inside of the Golgi or transfer to the site of GSL synthesis to initiate the synthesis of complex GSLs. Transferring to the GSL synthesis site is done with the help of a transport protein known as four phosphate adaptor protein 2 (FAPP2) while the flipping to the inside of the Golgi is made possible by the ABC transporter P-glycoprotein, also known as the multi-drug resistance 1 transporter (MDR1). GluCer is implicated in post-Golgi trafficking and drug resistance particularly to chemotherapeutic agents. For instance, a study demonstrated a correlation between cellular drug resistance and modifications in GluCer metabolism. In addition to their role as building blocks of biological membranes, glycosphingolipids have long attracted attention because of their supposed involvement in cell growth, differentiation, and formation of tumors. The production of GluCer from Cer was found to be important in the growth of neurons or brain cells. On the other hand, pharmacological inhibition of GluCer synthase is being considered a technique to avoid insulin resistance.

Ceramide-1-Phosphate

Ceramide-1-phosphate (C1P) is formed by the action of ceramide kinase (CK) enzymes on Cer. C1P carry ionic charge at neutral pH and contain two hydrophobic chains making

it relatively insoluble in aqueous environment. Thus, C1P reside in the organelle where it was formed and is unlikely to spontaneously flip-flop across membrane bilayers. C1P activate phospholipase A2 and is found, along with CK, to be a mediator of arachidonic acid released in cells in response to a protein called interleukin-1β (IL-1β) and a lipid-soluble molecule that transports calcium ions (Ca2+) across the bilayer, also known as calcium ionophore. C1P was also previously reported to encourage cell division (mitogenic) in fibroblasts, block apoptosis by inhibiting acid SMase in white blood cells within tissues (macrophages) and increase intracellular free calcium concentrations in thyroid cells.[59] C1P also has known roles in vesicular trafficking, cell survival, phagocytosis ("cell eating") and macrophage degranulation.

Second Messengers from Phosphatidylinositol

A general second messenger system mechanism can be broken down into four steps. First, the agonist activates a membrane-bound receptor. Second, the activated G-protein produces a primary effector. Third, the primary

Phosphatidylinositol bisphosphate (PIP2) Second Messenger Systems

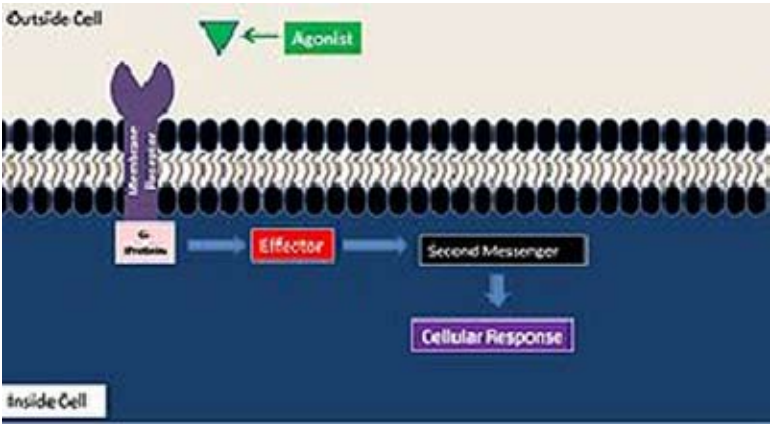


Fig. 16—Showing Cartoon of second messenger systems. Figure adapted From Barbraham Institute Mike Berridge.

effect stimulates the second messenger synthesis. Fourth, the second messenger activates a certain cellular process. The G-protein coupled receptors for the PIP2 messenger system produces two effectors, phospholipase C (PLC) and phosphoinositide 3-kinase (PI3K). PLC as an effector produces two different second messengers, inositol triphosphate (IP3) and Diacylglycerol (DAG). IP3 is soluble and diffuses freely into the cytoplasm. As a second messenger, it is recognised by the inositol triphosphate receptor (IP3R), a Ca^{2+} channel in the endoplasmic reticulum (ER) membrane, which stores intracellular Ca^{2+}. The binding of IP3 to IP3R releases Ca^{2+} from the ER into the normally Ca^{2+}-poor cytoplasm, which then triggers various events of Ca^{2+} signaling. Specifically in blood vessels, the increase in Ca^{2+} concentration from IP^3 releases nitric oxide, which then diffuses into the smooth muscle tissue and causes constrictions. DAG remains bound to the membrane by its fatty acid "tails" where it recruits and activates both conventional and novel members of the protein kinase C family. Thus, both IP3 and DAG contribute to activation of PKCs. Phosphoinositide 3-kinase (PI3K) as an effector phosphorylates phosphatidylinositol bisphosphate (PIP_2) to produce phosphatidylinositol (3, 4, 5)-trisphosphate (PIP_3). PIP_3 has been shown to activate protein kinase B, increase binding to extracellular proteins and ultimately enhance cell survival.

Activators of G-protein Coupled Receptors

Lysophosphatidic acid (LPA)

LPA is the result of phospholipase A2 action on phosphatidic acid. The SN-1 position can contain either an ester bond or an ether bond, with ether LPA being found at elevated levels in certain cancers. LPA binds the high-affinity G-protein coupled receptors LPA1, LPA2, and LPA3 (also known as EDG2, EDG4, and EDG7, respectively).

Sphingosine-1-phosphate (S1P)

S1P is present at high concentrations in plasma and secreted locally at elevated concentrations at sites of inflammation. It is formed by the regulated phosphorylation of sphingosine. It acts through five dedicated high-affinity G-protein coupled receptors, S1P1-S1P5. Interestingly, targeted deletion of S1P1 results in lethality in mice and deletion of S1P2 results in seizures and deafness. Additionally, a mere 3- to 5-fold elevation in serum S1P concentrations induces sudden cardiac death by an S1P3-receptor specific mechanism.

Platelet Activating Factor (PAF)

PAF is a potent activator of platelet aggregation, inflammation, and anaphylaxis. It is similar to the ubiquitous membrane phospholipid phosphatidylcholine except that it contains an acetyl-group in the SN-2 position and the SN-1 position contains an ether-linkage. PAF signals through a dedicated G-protein coupled receptor, PAFR and is inactivated by PAF acetylhydrolase.

Endocannabinoids

The endogenous cannabinoids, or endocannabinoids, are endogenous lipids that activate cannabinoid receptors. The first such lipid to be isolated was anandamide which is the arachidonoyl amide of ethanolamine. Anandamide is formed via enzymatic release from N-arachidonoyl phosphatidylethanolamine by enzymes which have not yet been delineated. It activates both the CB1 receptor, found primarily in the central nervous system, and the CB2 receptor which is found primarily in lymphocytes and the periphery. It is found at very low levels (nM) in most tissues and is inactivated by the fatty acid amide hydrolase. Subsequently, another endocannabinoid was isolated, 2-arachidonoylglycerol, which is produced when phospholipase C releases diacylglycerol which is then converted to 2-AG by diacylglycerol lipase. 2-AG

can also activate both cannabinoid receptors and is inactivated by monoacylglycerol lipase. It is present at approximately 100-times the concentration of anandamide in most tissues. Elevations in either of these lipids causes analgesia and anti-inflammation but the precise roles played by these two endocannabinoids are still vague and intensive research into their function, metabolism, and regulation is ongoing.

Prostaglandins

Prostaglandins are formed through oxidation of arachidonic acid by cyclooxygenases and other prostaglandin synthases. There are currently nine known G-protein coupled receptors (eicosanoid receptors) that largely mediate prostaglandin physiology (although some prostaglandins activate nuclear receptors, see below).

Retinol derivatives

Retinaldehyde is a retinol (vitamin A) derivative responsible for vision. It binds rhodopsin, a well-characterized GPCR that binds all-cis retinal in its inactive state. Upon photoisomerization by a photon the cis-retinal is converted to trans-retinal causing activation of rhodopsin which ultimately leads to depolarization of the neuron thereby enabling visual perception.

Activators of Nuclear Receptors

Steroid Hormones

This large and diverse class of steroids are biosynthesized from isoprenoids and structurally resemble cholesterol. Mammalian steroid hormones can be grouped into five groups by the receptors to which they bind: glucocorticoids, mineralocorticoids, androgens, estrogens, and progestagens.

Retinoic acid

Retinol (vitamin A) can be metabolized to retinoic acid

which activates nuclear receptors such as the RAR to control differentiation and proliferation of many types of cells during development.

Prostaglandins

The majority of prostaglandin signaling occurs via GPCRs (see above) although certain prostaglandins activate nuclear receptors in the PPAR family. (See article eicosanoid receptors for more information).

MAPK/ERK pathway is a chain of proteins in the cell that communicates a signal from a receptor on the surface of the cell to the DNA in the nucleus of the cell. The signal starts when a growth factor binds to the receptor on the cell surface and ends when the DNA in the nucleus expresses a protein and produces some change in the cell, such as cell division. The pathway includes many proteins, including MAPK (originally called ERK), which communicate by adding phosphate groups to a neighboring protein, which acts as an "on" or "off" switch. When one of the proteins in the pathway is mutated, it can be stuck in the "on" or "off" position, which is a necessary step in the development of many cancers. Components of the MAPK/ERK pathway were discovered when they were found in cancer cells. Drugs that reverse the "on" or "off" switch are being investigated as cancer treatments.

Coupling Cell Surface Receptors to G Proteins

Receptor-linked tyrosine kinases such as the epidermal growth factor receptor (EGFR) are activated by extracellular ligands. Binding of epidermal growth factor (EGF) to the EGFR activates the tyrosine kinase activity of the cytoplasmic domain of the receptor. The EGFR becomes phosphorylated on tyrosine residues. Docking proteins such as GRB2 contains an SH2 domain that binds to the phosphotyrosine residues of the activated receptor. GRB2 binds to the guanine nucleotide

exchange factor SOS by way of the two SH3 domains of GRB2. When the GRB2-SOS complex docks to phosphorylated EGFR, SOS becomes activated [3]. Activated SOS then promotes the removal of GDP from a member of the Ras subfamily (most notably H-Ras or K-Ras). Ras can then bind GTP and become active.Apart from EGFR, other cell surface receptors that can activate this pathway via GRB2 include Trk A/B, Fibroblast growth factor receptor (FGFR) and PDGFR.

Kinase Cascade

Activated Ras activates the protein kinase activity of RAF kinase. RAF kinase phosphorylates and activates MEK. MEK phosphorylates and activates a mitogen-activated protein kinase (MAPK).

RAF, MEK, and MAPK are all serine/threonine-selective protein kinases.

In the technical sense, RAF, MEK, and MAPK are all mitogen-activated kinases, as is MNK (see below). MAPK was originally called "extracellular signal-regulated kinases" (ERKs) and "microtubule-associated protein kinase" (MAPK). One of the first proteins known to be phosphorylated by ERK was a microtubule-associated protein (MAP). As discussed below, many additional targets for phosphorylation by MAPK were later found, and the protein was re-named "mitogen-activated protein kinase" (MAPK). The series of kinases from RAF to MEK to MAPK is an example of a protein kinase cascade. Such series of kinases provide opportunities for feedback regulation and signal amplification.

Regulation of Translation and Transcription

Three of the many proteins that are phosphorylated by MAPK are shown in the figure. One effect of MAPK activation is to alter the translation of mRNA to proteins. MAPK phosphorylates 40S ribosomal protein S6 kinase

(RSK). This activates RSK, which, in turn, phosphorylates ribosomal protein S6. Mitogen-activated protein kinases that phosphorylate ribosomal protein S6 were the first to be isolated MAPK regulates the activities of several transcription factors. MAPK can phosphorylate C-myc. MAPK phosphorylates and activates MNK, which, in turn, phosphorylates CREB. MAPK also regulates the transcription of the C-Fos gene. By altering the levels and activities of transcription factors, MAPK leads to altered transcription of genes that are important for the cell cycle.The 22q11, 1q42, and 19p13 genes are associated with schizophrenia, schizoaffective, bipolar, and migraines by affecting the ERK pathway.In simpler terms, the mitogen binds to the membrane ligand. This means that Ras (a GTPase) can swap its GDP for a GTP. It can now activate MAP3K (e.g., Raf), which activates MAP2K, which activates MAPK. MAPK can now activate a transcription factor, such as myc.

Clinical Significance

Uncontrolled growth is a necessary step for the development of all cancers. In many cancers (eg melanoma), a defect in the MAP/ERK pathway leads to that uncontrolled growth. Many compounds can inhibit steps in the MAP/ERK pathway, and therefore are potential drugs for treating cancer. eg. Hodgkin disease

The first drug licensed to act on this pathway is sorafenib - a Raf kinase inhibitor.

Other Raf inhibitors: SB590885, PLX4720, XL281, RAF265, PLX4032.

Some MEK inhibitors: XL518, CI-1040, PD035901, AZD6244, GSK1120212

Protein microarray analysis can be used to detect subtle changes in protein activity in signaling pathways.

JAK-STAT signaling pathway transmits information

from chemical signals outside the cell, through the cell membrane, and into gene promoters on the DNA in the cell nucleus, which causes DNA transcription and activity in the cell. The JAK-STAT system is a major signaling alternative to the second messenger system. The JAK-STAT system consists of three main components: a receptor, JAK and STAT.

JAK is short for Janus Kinase, and STAT is short for Signal Transducer and Activator of Transcription.

The receptor is activated by a signal from interferon, interleukin, growth factors, or other chemical messengers. This activates the kinase function of JAK, which autophosphorylates itself (phosphate groups act as "on" and "off" switches on proteins). The STAT protein then binds to the phosphorylated receptor. STAT is phosphorylated and translocates into the cell nucleus, where it binds to DNA and promotes transcription of genes responsive to STAT.

In mammals, there are seven STAT genes, and each one binds to a different DNA sequence. STAT binds to a DNA sequence called a promoter, which controls the expression of other DNA sequences. This affects basic cell functions, like cell growth, differentiation and death.

The JAK-STAT pathway is evolutionarily conserved, from slime molds and worms to mammals (but not fungi or plants). Disrupted or dysregulated JAK-STAT functionality (which is usually by inherited or acquired genetic defects) can result in immune deficiency syndromes and cancers.

Mechanism

JAKs, which have tyrosine kinase activity, bind to some cell surface cytokine receptors. The binding of the ligand to the receptor triggers activation of JAKs. With increased kinase activity, the phosphorylate tyrosine residues on the receptor and create sites for interaction with proteins that contain

phosphotyrosine-binding SH2 domains. STATs possessing SH2 domains capable of binding these phosphotyrosine residues are recruited to the receptors, and are themselves tyrosine-phosphorylated by JAKs. These phosphotyrosines then act as binding sites for SH2 domains of other STATs, mediating their dimerization. Different STATs form hetero- or homodimers. Activated STAT dimers accumulate in the cell nucleus and activate transcription of their target genes. STATs may also be tyrosine-phosphorylated directly by receptor tyrosine kinases, such as the epidermal growth factor receptor, as well as by non-receptor tyrosine kinases such as c-src.

The pathway is negatively regulated on multiple levels. Protein tyrosine phosphatases remove phosphates from cytokine receptors and activated STATs. More recently identified Suppressors of Cytokine Signaling (SOCS) inhibit STAT phosphorylation by binding and inhibiting JAKs or competing with STATs for phosphotyrosine binding sites on cytokine receptors. STATs are also negatively regulated by Protein Inhibitors of Activated STATs (PIAS), which act in the nucleus through several mechanisms. For example, PIAS1 and PIAS3 inhibit transcriptional activation by STAT1 and STAT3 respectively by binding and blocking access to the DNA sequences they recognize.

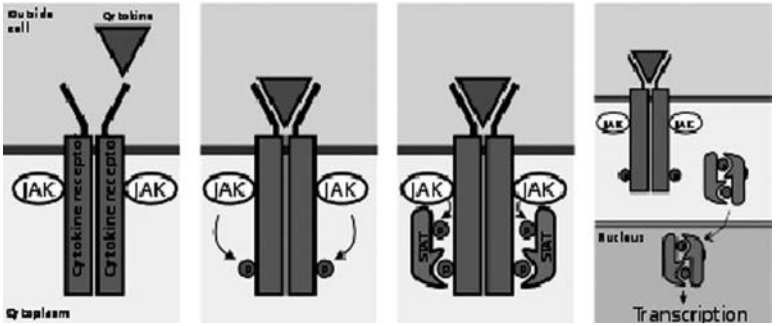


Fig. 17—Key steps of the JAK-STAT pathway

Chapter: Four

Exploration of Nano Systems and Devices

In the broadest sense, nano devices are the critical enablers that will allow mankind to exploit the ultimate technological capabilities of electronic, magnetic, mechanical, and biological systems. While the best examples of nano devices at present are clearly associated with the information technology industry, the potential for such devices is much broader. Nano devices will ultimately have an enormous impact on our ability to enhance energy conversion, control pollution, produce food, and improve human health and longevity.

Bio-MEMS is an abbreviation of biological micro-electromechanical systems and refers to a special class of micro-electromechanical systems (MEMS) where biological matter is manipulated to analyze and measure its activity under any class of scientific study. This class of devices belongs to one of the areas of development based on microtechnology. Among the applications based in Bio-MEMS are: biological and biomedical analysis and measurements and micro total analysis systems (TAS). One of the more popular approaches to Bio-MEMS as of late has been through biomimetics.

A *biosensor* is an analytical device for the detection of an analyte that combines a biological component with a physicochemical detector component.

It consists of 3 parts:

- the *sensitive biological element* (biological material (e.g. tissue, microorganisms, organelles, cell receptors, enzymes, antibodies, nucleic acids, etc), a biologically derived material or biomimic) The sensitive elements can be created by biological engineering.
- the *transducer* or the *detector element* (works in a physicochemical way; optical, piezoelectric, electrochemical, etc.) that transforms the signal resulting from the interaction of the analyte with the biological element into another signal (*i.e.*, transducers) that can be more easily measured and quantified;
- associated electronics or signal processors that are primarily responsible for the display of the results in a user-friendly way. This sometimes accounts for the most expensive part of the sensor device, however it is possible to generate a user friendly display that includes transducer and sensitive element (see Holographic Sensor).

A common example of a commercial biosensor is the blood glucose biosensor, which uses the enzyme glucose oxidase to break blood glucose down. In doing so it first oxidizes glucose and uses two electrons to reduce the FAD (a component of the enzyme) to FADH2. This in turn is oxidized by the electrode (accepting two electrons from the electrode) in a number of steps. The resulting current is a measure of the concentration of glucose. In this case, the electrode is the transducer and the enzyme is the biologically active component.

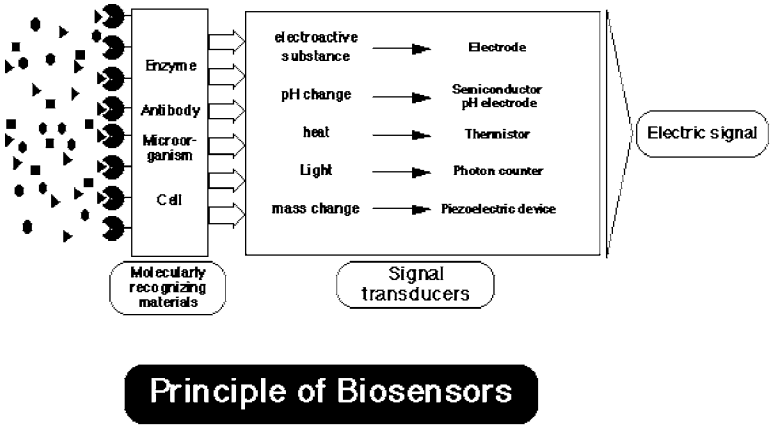


Fig. 30—Showing the Principle of Biosensors

Recently, arrays of many different detector molecules have been applied in so called electronic nose devices, where the pattern of response from the detectors is used to fingerprint a substance. Current commercial electronic noses, however, do not use biological elements.

A canary in a cage, as used by miners to warn of gas, could be considered a biosensor. Many of today's biosensor applications are similar, in that they use organisms which respond to toxic substances at much lower concentrations than humans can detect to warn of the presence of the toxin. Such devices can be used in environmental monitoring, trace gas detection and in water treatment facilities.

Principles of Detection

Photometric

Many optical biosensors based on the phenomenon of surface plasmon resonance are evanescent wave techniques. This utilises a property of gold and other materials; specifically

that a thin layer of gold on a high refractive index glass surface can absorb laser light, producing electron waves (surface plasmons) on the gold surface. This occurs only at a specific angle and wavelength of incident light and is highly dependent on the surface of the gold, such that binding of a target analyte to a receptor on the gold surface produces a measurable signal.

Surface plasmon resonance sensors operate using a sensor chip consisting of a plastic cassette supporting a glass plate, one side of which is coated with a microscopic layer of gold. This side contacts the optical detection apparatus of the instrument. The opposite side is then contacted with a microfluidic flow system. The contact with the flow system creates channels across which reagents can be passed in solution. This side of the glass sensor chip can be modified in a number of ways, to allow easy attachment of molecules of interest. Normally it is coated in carboxymethyl dextran or similar compound.

Light of a fixed wavelength is reflected off the gold side of the chip at the angle of total internal reflection, and detected inside the instrument. This induces the evanescent wave to penetrate through the glass plate and some distance into the liquid flowing over the surface.

The refractive index at the flow side of the chip surface has a direct influence on the behaviour of the light reflected off the gold side. Binding to the flow side of the chip has an effect on the refractive index and in this way biological interactions can be measured to a high degree of sensitivity with some sort of energy.

Other evanescent wave biosensors have been commercialised using waveguides where the propagation constant through the waveguide is changed by the absorption of molecules to the waveguide surface. One such example, Dual Polarisation Interferometry uses a buried waveguide as

a reference against which the change in propagation constant is measured. Other configurations such as the Mach-Zehnder have reference arms lithographically defined on a substrate. Higher levels of integration can be achieved using resonator geometries where the resonant frequency of a ring resonator changes when molecules are absorbed.

Other optical biosensors are mainly based on changes in absorbance or fluorescence of an appropriate indicator compound and do not need a total internal reflection geometry. For example, a fully operational prototype device detecting casein in milk has been fabricated. The device is based on detecting changes in absorption of a gold layer. A widely used research tool, the micro-array, can also be considered a biosensor. Biological biosensors often incorporate a genetically modified form of a native protein or enzyme. The protein is configured to detect a specific analyte and the ensuing signal is read by a detection instrument such as a fluorometer or luminometer. An example of a recently developed biosensor is one for detecting cytosolic concentration of the analyte cAMP (cyclic adenosine monophosphate), a second messenger involved in cellular signaling triggered by ligands interacting with receptors on the cell membrane. Similar systems have been created to study cellular responses to native ligands or xenobiotics (toxins or small molecule inhibitors). Such "assays" are commonly used in drug discovery development by pharmaceutical and biotechnology companies. Most cAMP assays in current use require lysis of the cells prior to measurement of cAMP. A live-cell biosensor for cAMP can be used in non-lysed cells with the additional advantage of multiple reads to study the kinetics of receptor response.

Electrochemical

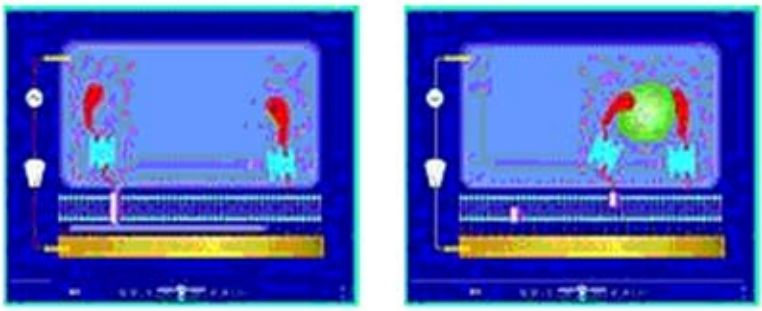
Electrochemical biosensors are normally based on enzymatic catalysis of a reaction that produces or consumes

electrons (such enzymes are rightly called redox enzymes). The sensor substrate usually contains three electrodes; a reference electrode, a working electrode and a sink electrode. An auxiliary electrode (also known as a counter electrode) may also be present as an ion source. The target analyte is involved in the reaction that takes place on the active electrode surface, and the ions produced create a potential which is subtracted from that of the reference electrode to give a signal. We can either measure the current (rate of flow of electrons is now proportional to the analyte concentration) at a fixed potential or the potential can be measured at zero current (this gives a logarithmic response). Note that potential of the working or active electrode is space charge sensitive and this is often used. Further, the label-free and direct electrical detection of small peptides and proteins is possible by their intrinsic charges using biofunctionalized ion-sensitive field-effect transistors.

Another example, the potentiometric biosensor, works contrary to the current understanding of its ability. Such biosensors are screenprinted, conducting polymer coated, open circuit potential biosensors based on conjugated polymers immunoassays. They have only two electrodes and are extremely sensitive and robust. They enable the detection of analytes at levels previously only achievable by HPLC and LC/MS and without rigorous sample preparation. The signal is produced by electrochemical and physical changes in the conducting polymer layer due to changes occurring at the surface of the sensor. Such changes can be attributed to ionic strength, pH, hydration and redox reactions, the latter due to the enzyme label turning over a substrate.

Ion Channel Switch

The use of ion channels has been shown to offer highly sensitive detection of target biological molecules. By imbedding the ion channels in supported or tethered bilayer



ICS - channel open ICS - channel closed

membranes (t-BLM) attached to a gold electrode, an electrical circuit is created. Capture molecules such as antibodies can be bound to the ion channel so that the binding of the target molecule controls the ion flow through the channel. This results in a measurable change in the electrical conduction which is proportional to the concentration of the target. An Ion Channel Switch (ICS) biosensor can be created using gramicidin, a dimeric peptide channel, in a tethered bilayer membrane. One peptide of gramicidin, with attached antibody, is mobile and one is fixed. Breaking the dimer stops the ionic current through the membrane. The magnitude of the change in electrical signal is greatly increased by separating the membrane from the metal surface using a hydrophilic spacer. Quantitative detection of an extensive class of target species, including proteins, bacteria, drug and toxins has been demonstrated using different membrane and capture configurations.

Others

Piezoelectric sensors utilise crystals which undergo an elastic deformation when an electrical potential is applied to

them. An alternating potential (A.C.) produces a standing wave in the crystal at a characteristic frequency. This frequency is highly dependent on the elastic properties of the crystal, such that if a crystal is coated with a biological recognition element the binding of a (large) target analyte to a receptor will produce a change in the resonance frequency, which gives a binding signal. In a mode that uses surface acoustic waves (SAW), the sensitivity is greatly increased. This is a specialised application of the Quartz crystal microbalance as a biosensor.

Thermometric and magnetic based biosensors are rare.

Applications

There are many potential applications of biosensors of various types. The main requirements for a biosensor approach to be valuable in terms of research and commercial applications are the identification of a target molecule, availability of a suitable biological recognition element, and the potential for disposable portable detection systems to be preferred to sensitive laboratory-based techniques in some situations. Some examples are given below:

- Glucose monitoring in diabetes patients—*historical market driver*
- Other medical health related targets
- Environmental applications e.g. the detection of pesticides and river water contaminants
- Remote sensing of airborne bacteria e.g. in counter-bioterrorist activities
- Detection of pathogens
- Determining levels of toxic substances before and after bioremediation
- Detection and determining of organophosphate
- Routine analytical measurement of folic acid, biotin,

vitamin B12 and pantothenic acid as an alternative to microbiological assay

- Determination of drug residues in food, such as antibiotics and growth promoters, particularly meat and honey.
- Drug discovery and evaluation of biological activity of new compounds.
- Protein engineering in biosensors http://www.springerlink.com/content/672p4l4l45xk02j2/
- Detection of toxic metabolites such as mycotoxins.

Glucose Monitoring

Commercially available gluocose monitors rely on amperometric sensing of glucose by means of glucose oxidase, which oxidises glucose producing hydrogen peroxide which is detected by the electrode. To overcome the limitation of ameperometric sensors, a flurry of research is present into novel sensing methods, such as fluorescent glucose biosensors.

Biosensors in Food Analysis

There are several applications of biosensors in food analysis. In food industry optic coated with antibodies are commonly used to detect pathogens and food toxins. The light system in these biosensors has been fluorescence, since this type of optical measurement can greatly amplify the signal.

A range of immuno- and ligand-binding assays for the detection and measurement of small molecules such as water-soluble vitamins and chemical contaminants (drug residues) such as sulfonamides and Beta-agonists have been developed for use on SPR based sensor systems, often adapted from existing ELISA or other immunological assay. These are in widespread use across the food industry.

Surface Attachment of the Biological Elements

An important part in a biosensor is to attach the biological elements (small molecules/protein/cells) to the surface of the sensor (be it metal, polymer or glass). The simplest way is to functionalize the surface in order to coat it with the biological elements. This can be done by polylysine, aminosilane, epoxysilane or nitrocellulose in the case of silicon chips/silica glass. Subsequently the bound biological agent may be for example fixed by Layer by layer deposition of alternatively charged polymer coatings.

Alternatively three dimensional lattices (hydrogel/ xerogel) can be used to chemically or physically entrap these (where by chemically entraped it is meant that the biological element is kept in place by a strong bond, while physically they are kept in place being unable to pass through the pores of the gel matrix). The most commonly used hydrogel is sol-gel, a glassy silica generated by polymerization of silicate monomers (added as tetra alkyl orthosilicates, such as TMOS or TEOS) in the presence of the biological elements (along with other stabilizing polymers, such as PEG) in the case of physical entrapment.

Another group of hydrogels, which set under conditions suitable for cells or protein, are acrylate hydrogel, which polymerize upon radical initiation. One type of radical initiator is a peroxide radical, typically generated by combining a persulfate with TEMED (Polyacrylamide gel are also commonly commonly used for protein electrophoresis), alternatively light can be used in combination with a photoinitiator, such as DMPA (2,2-dimethoxy-2-phenylacetophenone). Smart materials that mimic the biological components of a sensor can also be classified as biosensors using only the active or catalytic site or analogous configurations of a biomolecule.

ORNL's First Biosensor

The first biosensor developed at ORNL was intended for environmental monitoring. It used an antibody—a protein substance produced in the blood or tissues in response to a specific antigen, such as a bacterium or toxin normally foreign to the body. Antibodies destroy or weaken invading bacteria and neutralize organic poisons, forming the basis of immunity. Thus, ORNL's first biosensor was called an immunosensor.

In the mid-1980s, Tuan Vo-Dinh, Guy Griffin, and others in ORNL's Life Sciences Division were looking for a way to use light to detect cancer-causing agents in groundwater. So they attached to the end of an optical fi ber an antibody that reacts specifically with the carcinogen benzo(a)pyrene (BaP). The anti-BaP antibody on the end of the fiber was immersed in a sample of groundwater. The antibody was allowed to bind the BaP in the groun dwater sample.

The antibody-BaP reaction product gives off light if illuminated by light of the right wavelength. So the right light was aimed through the fiber into the groundwater sample. After 5 to 10 minutes, the reaction product fluoresced , and the fluorescence was transmitted back up the fiber and measured. These successful results, reported in 1987 by Vo-Dinh and colleagues, brought the group a 1987 R&D 100 Award from R&D magazine and initiated the group's development of a seri es of fiber-optic-based biosensors.

Antibodies can be produced against bacteria, against complex carbohydrates, and even against smaller organic molecules that may cause cancer. To Vo-Dinh's group, the possibilities for applications of immunosensors seemed almost limitless. Indeed the possible uses for biosensors are limited only by our imagination. After all, there are many different ways to combine chemistry, physics, and biology with an electronic detector.

One type of biosensor has only five components: a biological sensing element, a transducer, a signal conditioner, a data processor, and a signal generator. The essential component must produce a signal that is related to the concentration of a specific chemical or biological substance in complex systems. This component takes advantage of the ability of a biomolecule, such as an antibody or enzyme, to specifically recognize the target substance.

Light emissions from microspheres and bacteria are seen through a fluorescence microscope. Shown are red-fluorescing S. aureus bacteria bound to 6.5-μm spheres and one yellow orange-fluorescing E. coli bound to the larger 10-μm sphere.

In another approach to the use of immunosensors, microspheres of different sizes are labeled with antibodies that bind to different bacteria; thus, microspheres of one size have one particular antibody and microspheres of another s ize have a different antibody. The sizes of the microspheres are identified by their "morphological resonances" (shape-based light emissions when excited by a laser), and the bacteria that become bound are detected by the color of fluorescent dye with which they are stained. In a 1995 paper, Bill Whitten of ORNL's Chemical and Analytical Sciences Division suggests that up to 100 different types of bacteria can be identified simultaneous ly because the stained bacteria all would fluoresce at one wavelength of light and the diameters of the spheres could be illuminated at another wavelength.

Although all the spheres fluoresce when excited by one wavelength, the morphological resonances, which look like saw teeth superimposed on a fluorescence emission spectrum, can distinguish among diameters of many different-sized spheres. This approach satisfies one of today's challenges in biotechnology: multiplex biosensors to obtain more infor mation from one sample analysis.

Acoustic droplet ejection (ADE) uses a pulse of ultrasound to move low volumes of fluids (typically nanoliters or picoliters) without any physical contact. This technology focuses acoustic energy into a fluid sample in order to eject droplets as small as a picoliter.

ADE technology is a very gentle process, and it can be used to transfer proteins, high molecular weight DNA and live cells without damage or loss of viability. This feature makes the technology suitable for a wide variety of applications including proteomics and cell-based assays.

Ejection Mechanism

To eject a droplet, a transducer generates and transfers acoustic energy to a source well. When the acoustic energy is focused near the surface of the liquid, a mound of liquid is formed and a droplet is ejected. The diameter of the droplet scales inversely with the frequency of the acoustic energy—higher frequencies produce smaller droplets. Unlike other liquid transfer devices, no pipette tips, pin tools, or nozzles touch the source liquid or destination surfaces.

Liquid transfer methods that rely on droplet formation through an orifice, e.g., disposable tips or capillary nozzles, invariably lose precision as the transfer volume decreases. Touch less acoustic transfer provides a coefficient of variation (CVs) that is significantly lower than other techniques and is independent of volume at the levels tested. ADE shoots a droplet from a source well upward to an inverted receiving plate positioned above the source plate.

Liquids ejected from the source are captured by dry plates due to surface tension. For larger volumes, multiple droplets can be rapidly ejected from the source (typically 100 to 500 droplets/sec) to the destination with the coefficient of variation typically <4% over a volume range of two orders of magnitude.

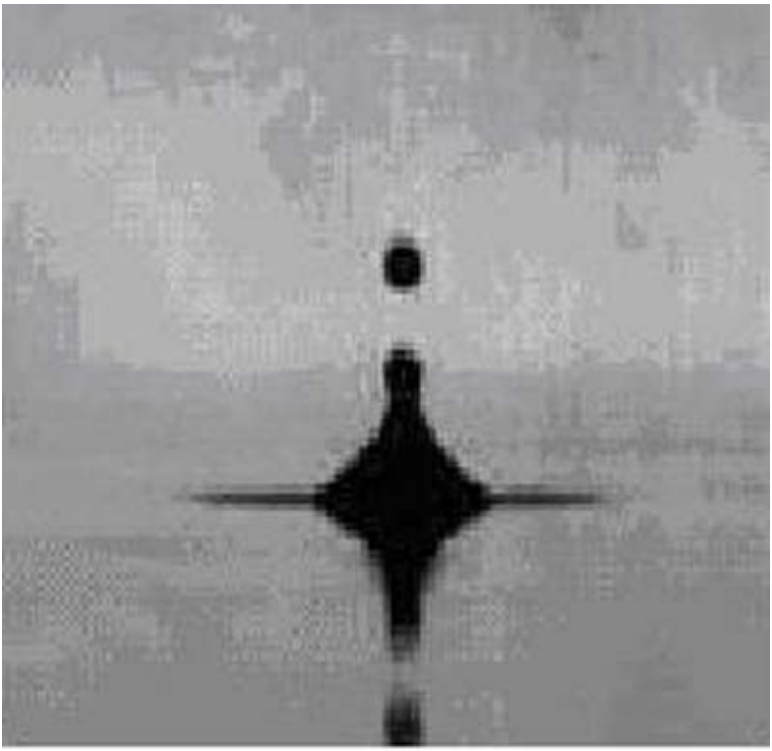


Fig. 41—Showing the Acoustic droplet system

Applications of Acoustic Transfer

The following applications are among those that can benefit from the features of acoustic droplet ejection:

- High throughput screening
- Microelectromechanical systems

Micro Pump

Although any kind of small pump is often referred to as micropump, a more accurate and up-to-date definition restricts this term to pumps with functional dimensions in the micrometre range. Such pumps are of special interest in

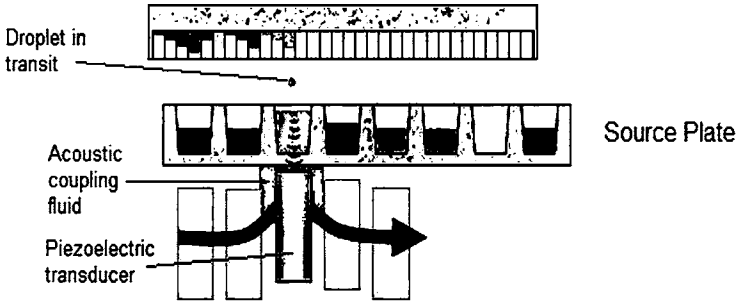


Fig 42—Showing the Application transfer Mechanism

microfluidic research, and have become available for industrial product integration in recent years. Their miniaturized overall size, potential cost and improved dosing accuracy compared to existing miniature pumps fuel the growing interest for this innovative kind of pump.

Types and Technology

In this sense, first true micropumps were reported on in 1975. However, the micropumps developed by Jan Smits and Harald Van Lintel in the early 1980's are considered to be the first genuine MEMS micropumps, and sparked the interest in shrinking the size of a fully functional pump to new dimensions. Within the microfluidic world physical laws change their appearance: As an example, volumetric forces, such as weight or inertia, often become negligible, whereas surface forces can dominate fluidical behaviour, especially when gas inclusion in liquids is present. With only a few exceptions, micropumps rely on micro-actuation principles, which can reasonably be scaled up only to a certain size.

Micropumps can be grouped into mechanical and non-mechanical devices: Mechanical systems contain moving parts, which are usually actuation and valve membranes or flaps. The driving force can be generated by utilizing piezoelectric, electrostatic, thermo-pneumatic, pneumatic

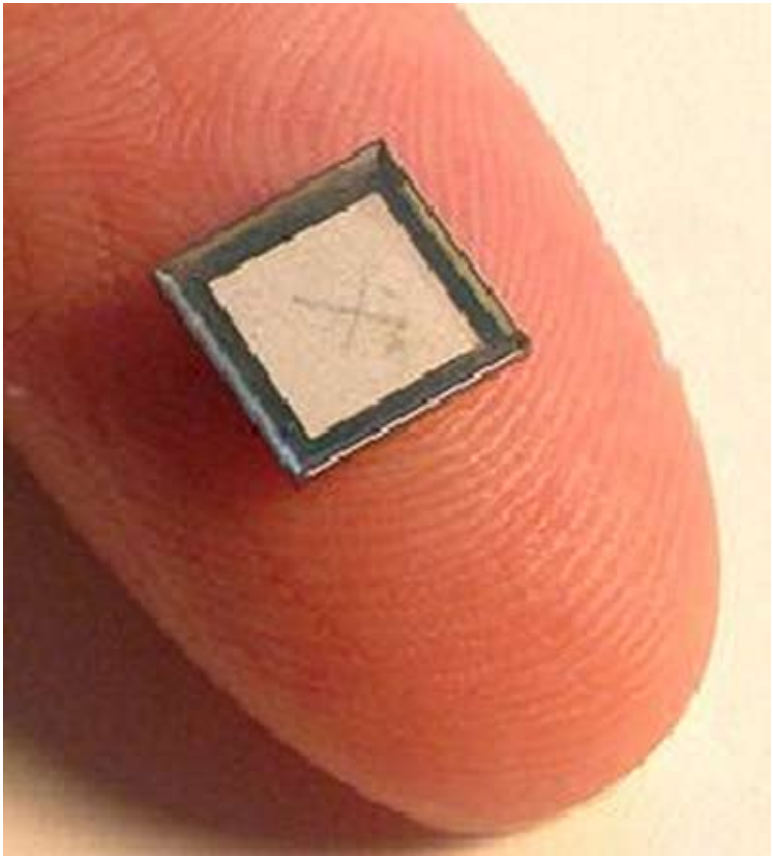


Fig. 42—Showing the micro pump

or magnetic effects. Non-mechanical pumps function with electro-hydrodynamic, electro-osmotic or ultrasonic flow generation, just to name a few of the actuation mechanisms that are currently studied.

Industrial Integration

Any kind of active microfluidic handling or analysis system (μTas, Lab-on-a-Chip System) requires some kind of micropump system. In addition, macro-fluidic systems which rely on miniature pumps might be reduced in size or enhanced

in their functionality by integrating a micropump. Emerging technologies, such as portable fuel cell applications will benefit when smaller yet more energy efficient pumps become available on the market. In 2003, the first commercial availability of a micropump was announced. Other companies have followed with their own pumps. All commercially available micropumps depend on piezoelectric actuation and incorporate passive check valves. Micropumps made of polymers appear to yield potentially low unit prices, while silicon micropumps prove to be the smallest pump devices in the world.

Lab on Chip

Experimental Information

Materials

The poly(dimethylsiloxane) (PDMS) prepolymer and curing agent were purchased from D.K scientific kolkata SU-8 negative photoresist and developer

The Bind-Silane solution consists of 0.8%(v,v) 3-(trimethoxysilyl)propyl methacrylate, 0.8%(v,v) acetic acid, pH 3.5 in ultra pure water. The gel salt bridge solution consists of 18%(w,v) acrylamide, 3%(w,v)N-methylenebisacrylamide and 3%(w,v) 2-dimethoxy-2- phenylacetophenone in isopropanol. Chemicals for the Bind-Silane, gel and sodium dodecyl sulfate (SDS) solutions were obtained from Sigma (St. Louis, MO). Fluorescein dye (F-1300) and the Live/Dead assay (L-3224) were obtained from Invitrogen. All solutions were prepared with filtered (0.2μm) ultra pure water (MilliQ, Millipore). For cell attachment a solution of fibronectin (F-4759, Sigma) in Delbecco's PBS was prepared. L15ex was prepared from the following recipe: 136mM NaCl, 5mM KCl, 1.6mM $MgSO_4$, 2.1 mM $MgCl_2$, 1.26mM $CaCl_2$, 1.3mM Na_2HPO_4, 0.36mMKH_2PO_4, 5mM Galactose, 6.7mM Pyruvate in filtered sterile ultra pure water.

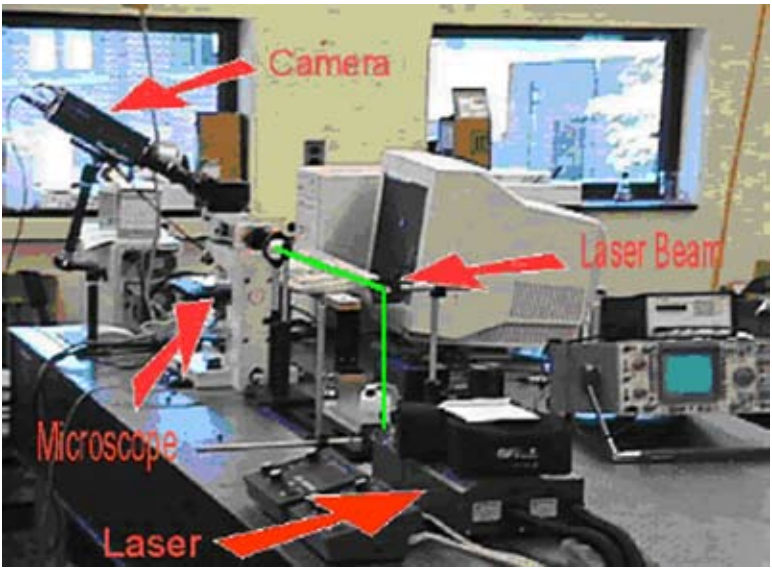


Fig. 31A—Showing author's Microfluidics lab

Fabrication of Electroosmotic Pumps and Toxicity Chip

Microfluidic devices were fabricated from PDMS using conventional soft-lithography techniques. 1, 2

Briefly, masks for fabricating SU-8 masters were designed in AutoCAD and printed in high resolution (20k dpi) on Mylar films (CAD/Art Services). The EO pumps and toxicity chip designswere patterned in SU-8 photoresist on separate 4" silicon wafers. Each master includes three replicasof the main design. The EO pump design contains multi-level structures of different height. The EOchannels were fabricated using SU-8 2001 (1.9μm thick); afterwards, SU-8 2025 was spun on top andthe gel regions were processed (95.4μm thick). The toxicity chip contains only one level of featuresthat were fabricated using SU-8 2025 (89.7μm thick). An optical profiler (Wyko NT1100, Veeco) was used to measure the dimensions of the relief features. PDMS was mixed in a ratio of 10:1 base to curing agent and degassed for 30 min in a vacuumchamber. The

liquid PDMS was poured onto each master and cured at 95°C for 12hr. Molds werethen cut out, trimmed and fluid access holes were punched. Microscope slides were ultrasoniccleaned in strong detergent for 10min, followed by 10min in ultra pure water, washed with ethanol, blow dried and baked at 200°C for 10min. The slide and PDMS mold were then air plasma treated

(PDC-001, Harrick Plasma) for 60s at 29.6W, then the two pieces were bonded.

Toxicity Chip Fabrication

After bonding, the gradient generator design was checked under a microscope to verify that the device was defect free. This step is important since the gradient profile depends strongly on the geometric design of the microfluidic channels. Then the chip was filled with ultra pure water, sealed with Kapton tape and placed under a UV sterilization lamp for 6hr. Completed chips were stored insterile bags until they were needed for toxicity experiments.

EO Pump Gel Salt Bridge Fabrication

Gel salt bridges were fabricated using a photolithography process as depicted in Fig. 31. Immediately after bonding, the EO pumps were filled with a freshly made Bind-Silane solution for 2hr. Plasma treatment is a precondition for the process to work effectively, since it exposes silanol groups,creating reaction sites for the Bind-Silane molecules. Bind-Silane is a bifunctional molecule thatparticipates in the polyacrylamide (PAA) reaction and creates a bond between the PAA gel and PDMS/glass, preventing leakage and increasing mechanical stability.3, 4 During the Bind-Silane treatment, the gel regions on the EO pumps were masked off with tape for theensuing photopolymerization. The gel solution was introduced into the chip by vacuum and the gel was polymerized by UV light

using a microscope with a 40X objective (GX-71, Olympus) and 100Wmercury burner with a 50% neutral density filter. The polymerization was monitored in situ with a CCD camera (sony ws30). The total time for processing all three EO pumps was10min.Afterwards, each EO pump was flushed by a series of solutions using a syringe pump to control the flow rate (11 Plus Dual Syringe Pump, Harvard Apparatus). The solutions were flushed from thefabrication inlets to the outlets of the EO pumps, so as to reduce the risk of clogging the narrow EOchannels. Each pump was flushed first with a 10% (v,v) solution of IPA to remove residual gelmonomer, then by a solution of 0.5% (w,v) NaOH to remove the Bind-Silane, and finally by ultra pure water to swell the gel. Each solution was flushed at a flow rate of 20μL min-1 for 10min. The fabrication access holes were then plugged with silicone sealant (732 RTV, Dow Corning) andreservoirs were attached. The EO pumps were filled with a 5mM Sodium Borate solution, sealed and left for two days before operating the pumps.

RTgill-W1 Cell Culture

The RTgill-W1 cell line was initiated from primary cultures of rainbow trout gill cells. Cells were maintained in non-vented flask at room temperature (22°C) with L-15 (Sigma Aldrich) media supplemented with 10% (v,v) fetal bovine serum, 100U mL-1 penicillin and 100μg mL-1 streptomycin. Cells used in these experiments have been passaged over 100 times.

Toxicity Testing Experimental Procedure

The protocol for performing the toxicity experiments is summarized in the flow chart presented in Fig. 32. Prior to performing experiments all items used with the cells were sterilized by autoclave. The cell chamber in the toxicity chip was flushed with a 1μg mL-1 fibronectin solution in Delbecco's PBS for 30min to promote cell attachment. The entire chip was then flushed with L-15 medium for an additional 30min.

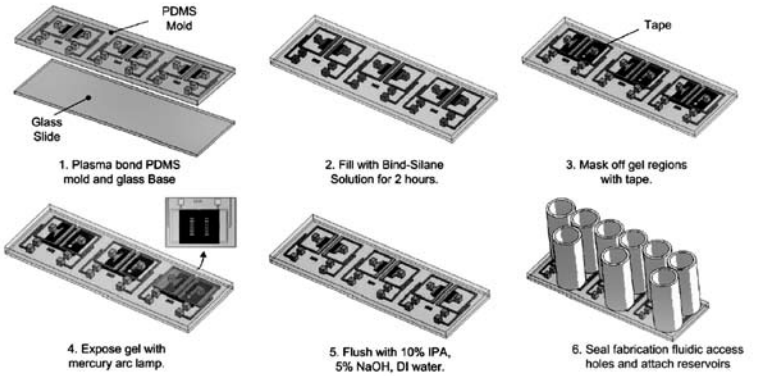


Fig. 31—Fabrication procedure for the EO pumps used in the toxicity studies. PDMS mold is bonded to a cleaned glass slide (1) and filled with Bind-Silane solution for 2hr (2). The chip is filled with the gel solution and the gel regions are taped off (3). The gels are photopolymerized by UV light on a microscope using a mercury arc lamp and CCD camera to monitor the process (4). The pumps are then flushed with solutions of 10% IPA, 0.5% NaOH and DI water (5). Finally, the fabrication access holes are sealed and reservoirs are attached (6) to complete the chip.

A cells-in-media suspension was made with an average cell density of 1.5x106 cells mL-1. The cell suspension was loaded into a syringe and injected into the cell chamber. The process was monitored under a phase contrast microscope to visualize cell movement. Once the cell density in the chamber was sufficient, the cells were allowed to attach in a static environment for 1hr, followed by perfusion with media for an additional 2hr.

During this period of time, solutions for the toxicity experiment were made. A serum free base solution of L15ex was used from which two types of solutions were prepared; one contained normal L15ex and the other L15ex with 50μg mL-1 of SDS. In addition, components of the Live/Dead cell assay kit were dissolved and vortexed in both solutions. The solutions were then loaded into the indirect pumping reservoirs and attached to EO pumps. Once the cell attachment period was

finished, the EO pumps and toxicity chip were connected. Earlier in the day the Q-P curves for each of the three pumps were obtained and used to determine the required current settings. First, the entire network was flushed with the L15ex using the one EO pump at a flow rate of 2μL min-1 for 20 min to remove unattached cells and debris. This also allowedfor the Live/Dead cell stain to incubate and highlight the cells prior to the experiment. Next the second pump with the SDS was set to 2μL min-1 and the gradient was allowed to form for 10min. At this point the toxicity experiment began and fluorescent images were recorded at 5min intervals for a period of 1hr using a fluorescence microscope (Eclipse E600, Nikon) with an FITC+Rhodamine filter set and camera (CoolPix E5400, Nikon). The investigation area was located 5mm downstream from the start of the cell chamber. The resulting images were analyzed to determine the fraction of dead cells by overlaying a rectangular mesh (2x3 mm) consisting of 20 rows, which equates to divisions of 2.5μg mL-1 of SDS.

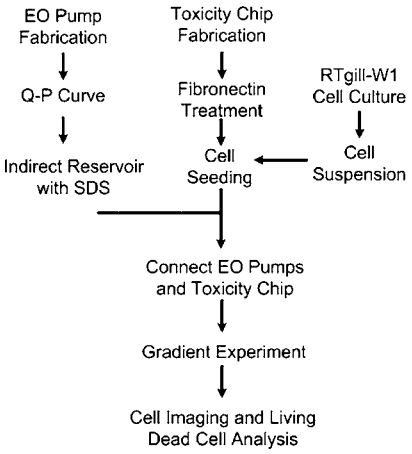


Fig. 32—Flow chart of the experimental procedure for toxicity study with EO pumps.

Chapter: Five

Nano Particles

Cornel Dots

By surrounding fluorescent dyes with a protective silica shell, Cornell University researchers have created fluorescent nanoparticles with possible applications in displays, biological imaging, optical computing, sensors and microarrays such as DNA chips. These are all applications for which quantum dots have been used or are being considered. But the new Cornell nanoparticles offer an appealing alternative because of their greater chemical inertness and reduced cost.

"People have done superb experiments with quantum dots that were not previously possible," says Ulrich Wiesner, Cornell associate professor of materials science and engineering. "Hopefully Cornell dots will serve the same purpose and offer new possibilities." There are also some interesting physics questions about how the new dots work, he adds. Since optical microscopes can't resolve individual molecules, and electron microscopes can't be used on living organisms, biologists often tag organic molecules with fluorescent dyes in order to track their movements through biological processes, such as the action of enzymes inside a living cell. While it can't see the molecules, an optical microscope can track the bright light

given off by the dye. Quantum dots -- which have been used for the same purpose -- are tiny particles of semiconductors such as cadmium selenide that behave as if they were individual atoms: They can absorb light energy, kicking their internal electrons up to higher energy levels, then release the energy by emitting light. A quantum dot fluoresces much more brightly than a dye molecule, making it a desirable marker. Cornell dots, also known as CU dots, are nanoparticles consisting of a core about 2.2 nanometers (nm) in diameter containing several dye molecules, surrounded by a protective silica shell, making the entire particle about 25 nm in diameter. The researchers

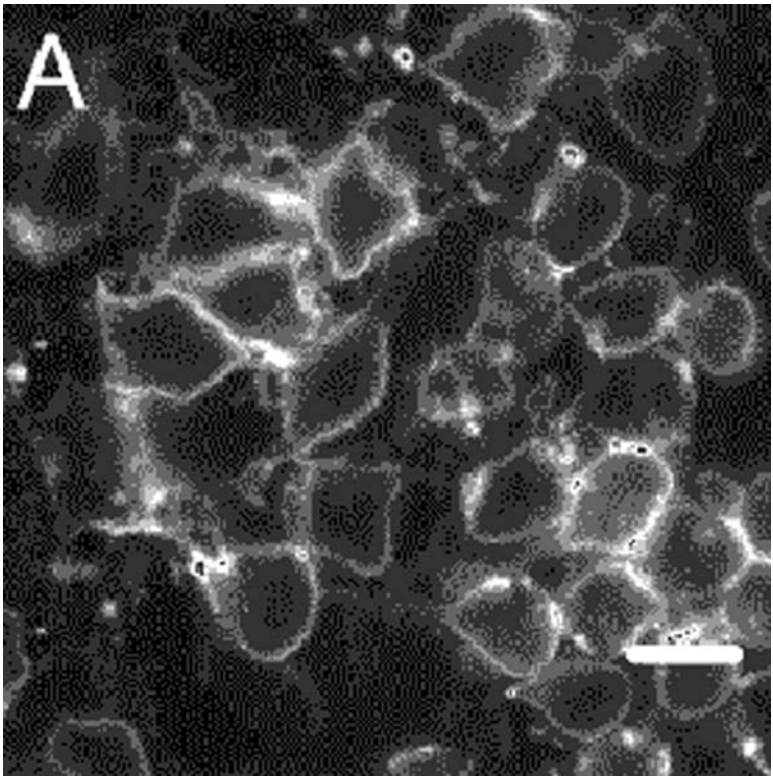


Fig. 33—CU dots bound to immunoglobin-G antibodies attach to the surface of leukemia cells, demonstrating a possible use in biological tagging.

call this a "core-shell architecture." (A nanometer is one-billionth of a meter, about three timesthe diameter of a silicon atom.) Like quantum dots, CU dots are many times brighter (20-30 times) than single dye molecules in solution and resist "photobleaching," a process by which dyes in solution rapidly lose their fluorescence. CU dots can be made with a wide variety of dyes, producing a large assortment of colors.

The manufacture of CU dots and early experiments with them are described in a paper, "Bright and Stable Core-Shell Fluorescent Silica Nanoparticles," in the journal Nano Letters (Vol. 5, No. 1) by Wiesner and his Cornell colleagues Hooisweng Ow, Daniel R. Larson, Mamta Srivastava, Barbara A. Baird and Watt W. Webb. Unlike quantum dots, CU dots are mostly chemically inert. The silica shell is silicon dioxide -- essentially glass. For use as biological markers, quantum dots are encased in a polymer shell -- a process that adds to their already high manufacturing cost. Quantum dots also contain heavy metals like cadmium that can leach through the polymer shell and disrupt the chemistry being observed. However, Wiesner says, "Silica is benign, cheap and easy to attach, and it is totally compatible with silicon manufacturing technology. That opens enormous possibilities in the life sciences and in information technology."

The Cornell researchers tested the use of CU dots as biological markers by attaching an antibody, immunoglobin E (IgE), and observing how this combination attached to cell receptors on leukemia mast cells. The dots also offer an intriguing physics question: Why do they fluoresce so brightly? In effect, the whole is brighter than the sum of its parts. "We have this enormous brightness, and we don't know exactly where it's coming from," Wiesner says. Several explanations have been offered. One is that the silicon shell protects the dye molecules from the solvent. A second is that dye molecules

floating free can lose energy by actions other than emitting photons, but in the packed core of the particle those other actions are diminished.

The dots were created by Ow, then Wiesner's graduate student. Webb, the S.B. Eckert Professor in Engineering, and Larson, a graduate student in applied and engineering physics now at Albert Einstein College of Medicine, studied their photophysical properties. Baird, director of the Cornell Nanobiotechnology Center, and Srivastava, a postdoctoral researcher, studied the dots as labels on living cells. The research was supported by the National Science Foundation, the state of New York and Phillip Morris USA. Quantum Dot Corp. supplied quantum dots used for comparison.

Brightly glowing nanoparticles known as "Cornell dots" are a safe, effective way to "light up" cancerous tumors so surgeons can find and remove them.

According to research at Memorial Sloan-Kettering Cancer Center (MSKCC), Cornell dots, also known as C dots, are biologically safe and stable and small enough to be easily transported across the body's structures and efficiently passed through the kidneys and out in urine.

A single dot consists of several dye molecules encased in a silica shell that can be as small as 5 nanometers in diameter (a nanometer is one-billionth of a meter, about three times the diameter of a silicon atom). The silica shell, essentially glass, is chemically inert. Coating the dots with polyethylene glycol, a process called PEGylation, further protects them from being recognized by the body as foreign substances, giving them more time to find targeted tumors. The outside of the shell can be coated with organic molecules that will attach to such desired targets as tumor surfaces or even locations within tumors. The cluster of dye molecules in a single dot fluoresces under near-infrared light much more brightly than single

Memorial Sloan Kettering Cancer Center

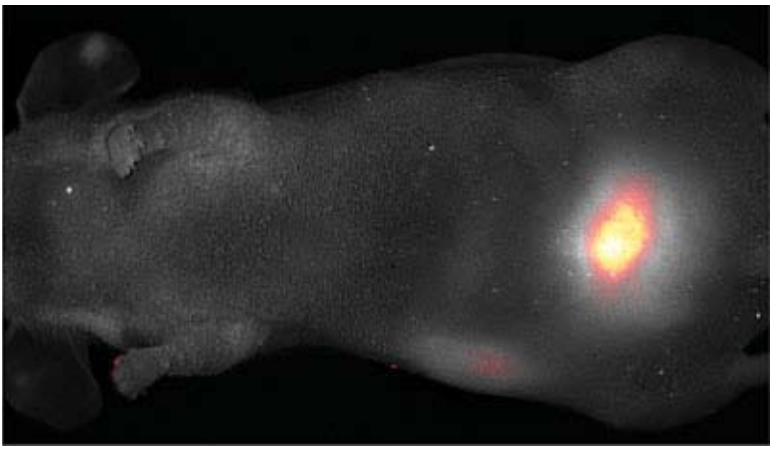


Fig 34—C dots fluoresce brightly enough to be seen through the skin of a mouse (faintly visible in this photo). Dots coated with polyethylene glycol have all reached the bladder in 45 minutes, demonstrating that C dots will be harmlessly excreted after they do their job.

dye molecules, and the fluorescence will identify malignant cells, showing a surgeon exactly what needs to be cut out and helping ensure that all malignant cells are found. According to MSKCC researchers, the technology also can show the extent of a tumor's blood vessels, cell death, treatment response and invasive or metastatic spread to lymph nodes and distant organs. Cornell dots were developed in 2005 by Hooisweng Ow (pronounced "Hoy-sweng-Oh"), then a graduate student working with Ulrich Wiesner, Spencer T. Olin Professor of Materials Science and Engineering at Cornell. Their refinements of the dot design and experiments in mice at MSKCC are reported in the January 2009 issue of the journal Nano Letters (Vol. 9 No. 1) by Wiesner, Dr. Michelle Bradbury, a physician-scientist specializing in molecular imaging and neuroradiology at MSKCC, and colleagues. "Highly sensitive and specific probes and molecular imaging strategies are critical to ensure the earliest possible detection of a tumor and

timely response to treatment," said Bradbury. "Our findings may now be translated to the investigation of tumor targeting and treatment in the clinic, with the goal of ultimately helping physicians to better tailor treatment to a patient's individual tumor."Since creating the Cornell dots, Wiesner, Ow and Kenneth Wang '77 have co-founded the company Hybrid Silica Technologies to commercialize the invention. The dots, Wiesner said, also have possible applications in displays, optical computing, sensors and such microarrays as DNA chips.The latest research was supported by the Clinical and Translation Science Center at Weill Cornell Medical College and the Cornell Nanobiotechnology Center. The original research was funded by the National Science Foundation, New York state and Phillip Morris USA.According to research at Memorial Sloan-Kettering Cancer Center (MSKCC), Cornell dots, also known as C dots, are biologically safe and stable and small enough to be easily transported across the body's structures and efficiently passed through the kidneys and out in urine. A single dot consists of several dye molecules encased in a silica shell that can be as small as 5 nanometers in diameter (a nanometer is one-billionth of a meter, about three times the diameter of a silicon atom). The silica shell, essentially glass, is chemically inert. Coating the dots with polyethylene glycol, a process called PEGylation, further protects them from being recognized by the body as foreign substances, giving them more time to find targeted tumors. The outside of the shell can be coated with organic molecules that will attach to such desired targets as tumor surfaces or even locations within tumors. The cluster of dye molecules in a single dot fluoresces under near-infrared light much more brightly than single dye molecules, and the fluorescence will identify malignant cells, showing a surgeon exactly what needs to be cut out and helping ensure that all malignant cells are found. According to MSKCC researchers, the technology also

can show the extent of a tumor's blood vessels, cell death, treatment response and invasive or metastatic spread to lymph nodes and distant organs. Cornell dots were developed in 2005 by Hooisweng Ow (pronounced "Hoy-sweng-Oh"), then a graduate student working with Ulrich Wiesner, Cornell associate professor of materials science and engineering. Their refinements of the dot design and experiments in mice at MSKCC are reported in the January 2009 issue of the journal Nano Letters (Vol. 9 No. 1) by Wiesner, Dr. Michelle Bradbury, a physician-scientist specializing in molecular imaging and neuroradiology at MSKCC, and colleagues. "Highly sensitive and specific probes and molecular imaging strategies are critical to ensure the earliest possible detection of a tumor and timely response to treatment," said Bradbury. "Our findings may now be translated to the investigation of tumor targeting and treatment in the clinic, with the goal of ultimately helping physicians to better tailor treatment to a patient's individual tumor."Since creating the Cornell dots, Wiesner, Ow and Kenneth Wang '77 have co-founded the company Hybrid Silica Technologies to commercialize the invention. The dots, Wiesner said, also have possible applications in displays, optical computing, sensors and such microarrays as DNA chips.The latest research was supported by the Clinical and Translation Science Center at Weill Cornell Medical College and the Cornell Nanobiotechnology Center. The original research was funded by the National Science Foundation, New York state and Phillip Morris USA.